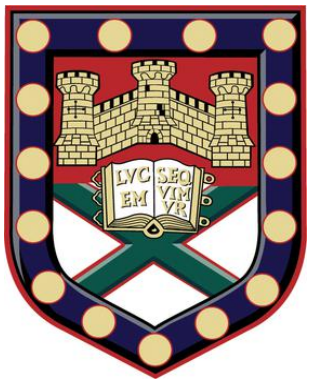
A Systems Biology Approach to Investigating Host-Pathogen Interactions in Infection with *Burkholderia pseudomallei*



Submitted by Deniz Cizmeci to the University of Exeter

as a thesis for the degree of

Doctor of Philosophy in Mathematics

In February 2017

ABSTRACT

This thesis applies systems approaches in order better to understand host-pathogen interactions in infectious diseases; it focuses on the intracellular bacterium *Burkholderia pseudomallei*, the causative agent of the human disease melioidosis. Little is known about the epigenetic changes in host cells during infection. This study assesses genome-wide patterns of the epigenetic marker DNA methylation in host cells following infection with *B. pseudomallei*. The studies of this thesis concern the infection of human macrophage-like U937 cells with *B. pseudomallei* and the DNA methylation levels were measured during the early stages of infection. Analyses reveal significant changes in infected cells (compared to uninfected controls) at multiple locations in the host DNA. Most of the methylation changes in infected cells are losses rather than gains in methylation. Five different differential methylation patterns (constant, early, late, transient, and oscillatory) are identified. Differentially methylated sites mapped to genes that may affect virulence, e.g. genes involved in actin regulation, immune response, inflammatory response, and nitric oxide generation. The thesis also measures whole blood DNA methylation profiles of patients diagnosed with melioidosis in order to test the potential role of host DNA methylation in melioidosis. The results demonstrate that patients with melioidosis are separated from healthy subjects by their distinct methylation profiles. The differentially methylated regions reported here can potentially be used as biomarkers for classification and prognostication of infectious diseases. In addition to exploring the changes to the host, a comprehensive understanding of the pathogen interference and the search for countermeasures requires a framework that assesses how the host changes

the pathogen metabolically. In this thesis, to understand the role of trehalose pathway in virulence, computational models were constructed by integrating kinetic information, genomics data and literature surveys. Existing kinetic models of the trehalose pathway were implemented and extended allowing for the *in silico* investigation of the trehalose mutant. Further, metabolic networks of *B. pseudomallei* were analysed at the genome scale to identify molecular links between trehalose and metabolic pathways such as glycolysis. The genome-scale reconstruction of the *B. pseudomallei* metabolic network was used to simulate growth under different conditions and predict the effects of gene knockouts.

This thesis not only expands the existing knowledge about *B. pseudomallei* infection, the novel approaches employed here will stimulate a wider understanding of the applications of systems biology to host-pathogen research and defence needs.

Acknowledgements

I have been more than fortunate to have amazing people in my life who for me define words like trust, encouragement, support, love, patience. I am grateful for the opportunity to thank them, although words could not express the full extent of my gratitude.

I would like to start by thanking Professor Orkun Soyer, Professor Rick Titball, Professor Jo Prior, and Dr Ozgur Akman for giving me the opportunity to undertake this research project. I am grateful for their guidance, support, and encouragement. It has been an honour working with them.

I would also like to thank Dr. Emma Dempster, Dr. Olivia Champion, Dr. Sariqa Wagley for their invaluable role in my project.

Special thanks to all my incredible friends who made Exeter home, the most amazing people! Tom, Callan, Silvia, Nathan, Will, Alex, Ed, Juliette, Jani, JC, Melek, Pam, David.

Above All, Andrea, I feel lucky everyday to have you in my life.

Last but not least I wish to express my gratitude to my parents, Guler and Hamdi, and my brother, Hakan, for their boundless love, support and sacrifices, for always believing in me, for providing a solid ground to stand on, and for providing me with an inherent sense of security throughout my life.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

COG	Clusters of Orthologous Groups
CDD	Conserved Domain Database
DAPI	4',6-diamidino-2-phenylindole
ddNTP	dideoxynucleotide
DMP	Differentially methylated position
DMR	Differentially methylated region
DNMT	DNA methyltransferases
EFM	Elementary Flux Mode analysis
FCPT	Fisher's combined probability test
FBA	Flux balance analysis
FBS	foetal bovine serum
GO	Gene ontology
LB	Luria-Bertani
MBP	Methyl-CpG-binding proteins
NGS	Next-generation sequencing
OR	odds ratio
PATRIC	Pathosystems Resource Integration Center
PMA	phorbol 12-myristate 13-acetate
PFS	phosphate buffered saline
PCA	Principle component analysis
RFP	red fluorescent protein
T2D	Type 2 diabetes
WGCNA	Weighted Gene Correlation Network Analysis

Publications Arising from this Thesis

The results presented in Chapter 2 have been published as a journal article. The accepted manuscript is presented in Appendix.

Cizmeci D, Dempster EL, Champion OL, Wagley S, Akman OE, Prior JL, Soyer OS, Mill J, Titball RW. Mapping epigenetic changes to the host cell genome induced by *Burkholderia pseudomallei* reveals pathogen-specific and pathogen-generic signatures of infection. Scientific Reports. 2016;6.

The data generated in Chapter 3 will be included in a future publication on the DNA methylation changes in patients with septicemic melioidosis.

The mathematical models in Chapter 4 provide future directions for experiments on the trehalose metabolism in *B. pseudomallei*. These models will potentially be published when validated with experimental data.

Declarations

Unless otherwise stated, the results and data presented in this thesis were solely my work.

- The cell line infection experiments presented in Chapter 2 were performed by Dr. Olivia L. Champion (Experiment 1) and Dr. Sariqa Wagley (Experiment 2) at the University of Exeter. DNA and RNA isolations with Qiagen AllPrep method in Chapter 2 were carried out by Dr. Olivia L. Champion and Dr. Sariqa Wagley.
- The mouse lung tissues used in Chapter 2 were provided by Prof. Gregory Bancroft from London School of Hygiene & Tropical Medicine.
- DNA and RNA isolations using the Trizol method in Chapter 2 were done by me with assistance from Dr. Olivia L. Champion and Dr. Sariqa Wagley.
- Whole blood samples used in Chapter 3 were provided by Prof. Ganjana Lertmemongkolchai from Khon Kaen University.
- The samples used in Chapter 2 and Chapter 3 were processed on the DNA methylation array by Dr. Emma L. Dempster at the University of Exeter Medical School. PCR analyses in Chapter 2 were conducted by me with assistance from Dr. Emma L. Dempster.
- Probe level functional analysis in Chapter 2 and regional analysis using Brown's method in Chapter 3 were adapted from scripts developed by Dr. Eilis Hannon at the University of Exeter Medical School.
- Benjamin P. Lee at the University of Exeter Medical School assisted with Q-PCR analysis presented in Chapter 2.
- The genome scale *Burkholderia pseudomallei* K96243 SEED272560_3 model used in Chapter 4 was constructed by Kalesh Sasidharan at the University of Warwick.

CHAPTER 1. INTRODUCTION

1.1 Infectious diseases and host-pathogen interactions

Infectious diseases are a major problem worldwide. Diverse groups of pathogens—bacteria, viruses, fungi, protozoa—are able to produce infections ranging from mild to severe conditions. Human diseases such as tuberculosis, malaria, HIV and neglected tropical diseases kill more than 4 million people every year [1]. The emergence of drug resistance [2] and the lack of effective vaccines [3] have intensified the need for novel approaches to understanding the biology driving infection and to accelerate the development of therapies.

A major current focus in infectious disease research is host-pathogen interactions. Pathogens adopt various strategies to survive and replicate within their hosts, often subverting host cellular pathways, while the hosts initiate responses to control pathogen activity [4]. Bacterial pathogens usually inject specialized proteins—pathogenic factors—directly into host cells. These factors can target crucial intracellular pathways at multiple points. Hence, a more comprehensive picture of the functional consequences can be achieved when pathogen interference is evaluated in the context of the host networks it operates on.

1.2 Systems analysis of host pathogen interactions

Traditional approaches focus on single genes or gene-for-gene interactions, which are likely to play a role in the infection process. However, cellular functions are hardly ever dictated by single genes acting alone. Instead, the concerted action of multiple components comprises cellular processes that can give rise to biological behaviour [4]. Thus, gene-for-gene interaction models

are insufficient in capturing the whole picture of the function and evolution of pathogen interference. Systems approaches, by which genome-wide information is screened and integrated, is vital for understanding the complex mechanisms of disease. Systems biology data at the organism or cellular scale is often produced through system-wide (referred to as "–omic") technologies and analysed by mathematical modelling. Sections 1.2.1 and Section 1.2.2 review high-throughput screening technologies and the perspectives these offer in terms of host-pathogen interactions. Section 1.2.3 gives an overview of mathematical modelling methods and their application to host-pathogen research.

1.2.1 -Omic Technologies

The postgenomic era has provided a huge amount of knowledge on the genetic makeup of organisms. Systematic large-scale analyses allow faster and high-throughput screening to test hypotheses at various stages of biological processes [5]. Omic technologies—from whole-genome sequencing, to transcriptomics, proteomics and metabolomics—have been invaluable in deciphering the genetic underpinnings of complex biological traits [6–8].

Next-generation sequencing (NGS) technologies have allowed the rapid and cheap sequencing of DNA and RNA to study the genomics of infectious diseases. The availability of whole genome sequences for thousands of pathogens has advanced the understanding of pathogenesis. For example, researchers identified genomic signatures pointing to the emergence of drug resistance [9] or have located the outbreak of a particular disease [10,11].

Gene expression is central to biological function, meaning that transcriptomics has been instrumental in infectious disease research [12].

Transcriptomics measures the expression levels of mRNAs, which can vary with external conditions. Such changes occur when the host and the pathogen meet [13,14]. The transcriptional response of the host or the pathogen during infection has been evaluated to assess the dynamics of infection, to characterise virulence factors, to identify biomarkers, and to study host immunity [12-14]. More sophisticated approaches profiled expression levels of the host and pathogen simultaneously [15,16]. Microarray and sequencing technologies are used for transcriptome-wide gene expression profiling. Microarrays determine the abundance of nucleic acids by hybridization to probes on microchips. Microarray chips have been mostly developed for large transcriptomes (i.e. humans, mice). While microarrays can efficiently analyse large numbers of genes, these genes are targets of predetermined probes. In contrast, sequencing techniques for transciptomics (RNA-seq) offer the identification of novel transcripts and are preferred for discovery analyses in non-model organisms such as pathogenic microorganisms [15]. In addition to transcriptomics, other system-level approaches, which inspect gene-protein interactions and metabolic pathways, have been employed to decipher the virulence of various pathogens [16–19].

The precise temporal and spatial control of gene expression requires the concerted action of transcription factors as well as epigenetic modulators. Epigenetic modulators affect how genes are read without changes in the DNA sequence [20]. There are many types of epigenetic processes including DNA methylation, histone modifications, and non-coding RNA. These processes can alter chromatin structure and control gene expression [21]. In the drive towards a complete understanding of infectious diseases, it is crucial to consider the role of epigenetics. Despite the significant advances in epigenetic technologies and

evidence supporting epigenetic modulation of host-pathogen interactions, little is known about the epigenetic changes in host cells during infectious diseases. Chapters 2 and 3 of this thesis present evidence for the contribution of epigenetic changes to host response during infection by focusing on DNA methylation. As background for these chapters, Section 1.4 provides an overview of epigenetic mechanisms and how DNA methylation may contribute to host-initiated and/or pathogen-induced responses.

1.2.2 Interpreting high-throughput data and network analyses

The ample information obtained through high-throughput technologies brings up the next challenge: to integrate this knowledge into multi-scale models of host-pathogen interaction pathways and to provide a holistic understanding of physiological and disease states. To meet these challenges, computational tools have been developed. These bioinformatics methods can leverage the huge amounts of data for valuable analysis. In particular, statistical tests are being employed to determine the significance of the results (e.g. changes in gene expression) drawn from analyses [22]. High-throughput data are often high dimensional, especially in cases of complex experimental setups (e.g. comparisons between multiple groups of samples, time point measurements, etc.). Clustering analyses reduce the complexity by grouping variables based on their similarity in the data [23,24]. Gene ontology [25] and pathway analyses [26] are frequently applied to provide a biological context to the large number of differentially regulated genes obtained from genomics data.

1.2.3 *In silico* modelling and simulation

While bioinformatics tools organize and aid the interpretation of these data, systems biology approaches adopt methods to integrate the multiple types of information into a mathematical format to infer properties about that system [27]. Thus, beyond the analysis of network organization, the available information can be merged into dynamic, explanatory and predictive models [28]. The calculation of network characteristics provides explanations for biological phenomena. Simulation of different scenarios enables testing of hypotheses, which may not be possible or feasible in laboratory experiments.

Such mathematical models have been invaluable in studying the role of metabolism in host pathogen interactions [29,30]. Mathematical models of metabolism have been developed using two major modelling methodologies. Metabolic models contain either a few reactions described to a high level of kinetic detail (kinetic models) or a large set of reactions with little or no kinetic information (constraint-based models) [31].

Kinetic models allow for dynamic simulation and control analysis. To build a kinetic model, rate laws need to be assigned to all reactions and the kinetic constants in each reaction must be determined. Extensive experimental data is required to characterise the mechanics of a kinetic model. For this reason, kinetic models are often available merely for central pathways of well-studied model organisms, such as *Escherichia coli* or *Saccharomyces cerevisiae* [32,33]. Nonetheless these models offer an indispensible insight into the dynamic regulation of core metabolism and reveal underlying mechanisms of phenotypes (e.g. adaptation to carbon sources) that can be applied to a wider range of organisms [32,34,35].

On the other hand, the increasing number of fully sequenced genomes has allowed the construction of constraint-based models for a wide range of organisms [36–38]. Constraint-based models only account for the stoichiometry and directionality of reactions, which can be derived from genome annotations. These metabolic reconstructions can serve to evaluate the biosynthetic routes for production of toxic materials by pathogens within the host cells or to the survival capacities of the pathogen within the host-cell nutrient environment. Flux balance analysis (FBA) [39] has been the leading constraint-based tool. FBA calculates steady-state metabolic fluxes and can predict the growth rate of an organism. FBA can simulate the deletion of genes and survey the essentiality of reactions in various pathogens including bacterial species [40]. Metabolic reconstruction of pathogens such as *Mycobacterium tuberculosis* [41], *Salmonella typhimurium* [42], and *Francisella tularensis* [43] offered cost-effective pipelines to search for potential targets for medical interventions.

1.3 Aims and Overview of the thesis

This thesis applies systems approaches to enhance the understanding of host-pathogen interactions in infectious disease. Within the host context, potential host targets of either pathogen-mediated or pathogen-initiated epigenetic changes in the host were identified. As there were very limited reports in the literature on the epigenetic changes after infection, this work uncovers novel mechanisms of virulence of pathogens and the host response. In addition to exploring the changes to the host, this thesis assesses how the pathogen changes metabolically within the host environment. This multifaceted view on pathogen interference will stimulate new directions for research into novel disease interventions. The particular focus of this thesis is the intracellular bacterium *Burkholderia pseudomallei*, the causative agent of melioidosis; however, many of the approaches can be applied to other diseases and pathogens.

The key objectives of this thesis are: (i) to identify genome-wide changes in host DNA post infection; (ii) to map temporal changes over the course of infection; (iii) to understand the functional significance of the interference with the host; and (iv) to understand the role of bacterial metabolism in virulence.

The organisation of the thesis is as follows:

Chapter 1 provides a literature review on host parasite interactions in infectious diseases and the system approaches used. Section 1.4 provides a background on epigenetics and DNA methylation, Section 1.5 gives motivation for interest in trehalose metabolism and Section 1.6 introduces *B. pseudomallei*.

Chapter 2 reports a genome wide analysis of changes in the DNA methylation state of host cell genes following infection with the intracellular

pathogen *B. pseudomallei* in an *in vitro* infection model. A human macrophage cell line was used.

Chapter 3 identifies DNA methylation differences associated with melioidosis, an infectious disease caused by *B. pseudomallei*. To this end, whole blood DNA methylation levels of patients diagnosed with septicemic melioidosis and healthy subjects were measured.

Chapter 4 focuses on the metabolic network of *B. pseudomallei* and mathematical modelling approaches to understanding the role of metabolism in virulence. Particular emphasis is given to the trehalose pathway. The importance of the trehalose pathway in bacterial virulence is reviewed in Section 4.1.

Chapter 5 draws the conclusions of this study and provides suggestions for further work.

1.4. Epigenetics in host-pathogen interactions

1.4.1 Epigenetics

Epigenetics is defined as the study of heritable phenotypes that are associated with changes in gene expression without changes in the DNA sequence [20,44]. Epigenetics refers to the extra layer of information, on top of the genetic code, that allows development and differentiation of diverse functions from the same set of instructions [45]. Epigenetic processes can alter DNA accessibility and chromatin structure, thereby regulating gene expression and activity [46]. These processes are highly dynamic and exhibit tissue-specific patterns [47,48]. Epigenetic information can be retained and passed on to offspring [49,50]. Modifications occurring through these epigenetic processes can dictate differentiation of different cell types from identical DNA [51]. Epigenetic modifications also modulate patterns of gene expression in response to environmental stimuli [52,53]. Such modifications are therefore vital in interpreting the genome under physiological and pathological conditions [54]. Profiling epigenetic modifications of the entire genome—the epigenome in a given cell—may serve as biomarkers for diagnosis and therapeutic tools [55–57].

Epigenetic mechanisms include DNA methylation, histone modifications and non-coding RNA. Histones are special proteins, which wrap eukaryotic DNA into chromatin structure [58]. Modifications of histones (e.g acetylation, methylation, phosphorylation, etc.) can result in alterations in the chromatin architecture inhibiting the access of the transcriptional machinery to DNA and ultimately affecting gene expression [21]. Furthermore, non-coding RNA—RNA that does not encode a protein—can be involved in controlling chromatin structure and gene expression [59]. This thesis focuses on DNA methylation,

which is described in the next section.

1.4.2 DNA methylation

In eukaryotes, DNA methylation involves the transfer of a methyl group to the C-5 position of the cytosine ring by DNA methyltransferases (DNMTs) and usually occurs at CpG sites (Figure 1.1) [60]. CpG sites are regions of DNA where a cytosine nucleotide occurs next to a guanine nucleotide in the DNA sequence. CpG islands, regions where high frequencies of CpG sites are clustered, are often found close to the promoters of genes, where the transcription machinery binds [61] and are typically unmethylated [45].

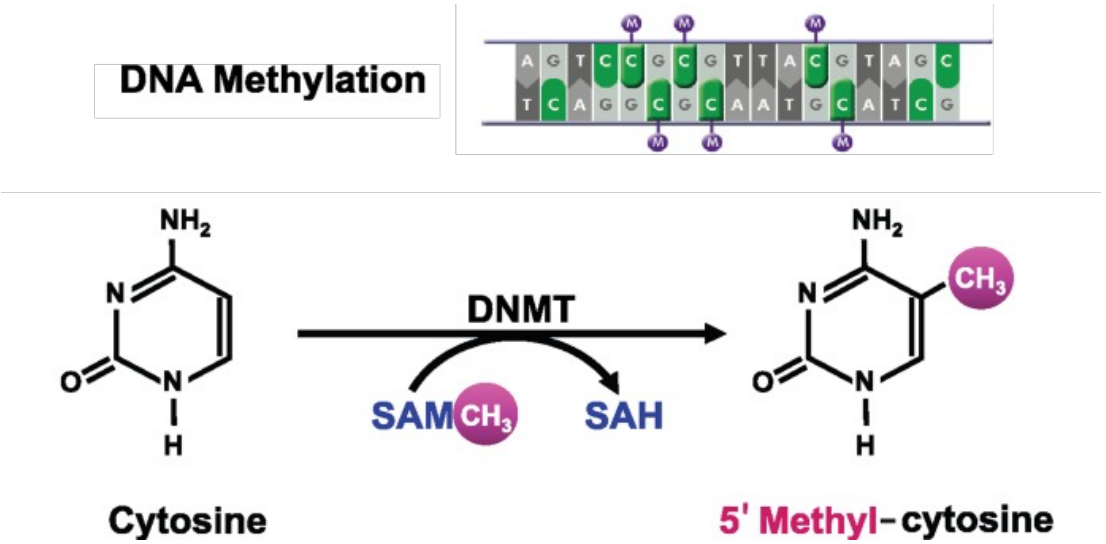


Figure 1.1. DNA methylation process. DNA methylation often occurs at cytosine nucleotides that are followed by a guanine nucleotide (CpG site) and involves the conversion of cytosine to 5'methyl-cytosine by DNA methyltransferase (DNMT). S-adenosylmethionine (SAM) is used as a methyl donor and converted to S-adenosylhomocysteine (SAH). Figure and legend adapted from [62].

DNA demethylation can occur passively or actively: passive DNA demethylation can occur as a consequence of cell growth and DNA replication in the absence of DNA methyltransferases [63]; active DNA demethylation

involves enzymatic removal of methylation (oxidation mediated by TET proteins [64]) via intermediates [65–67] as shown in Figure 1.2.

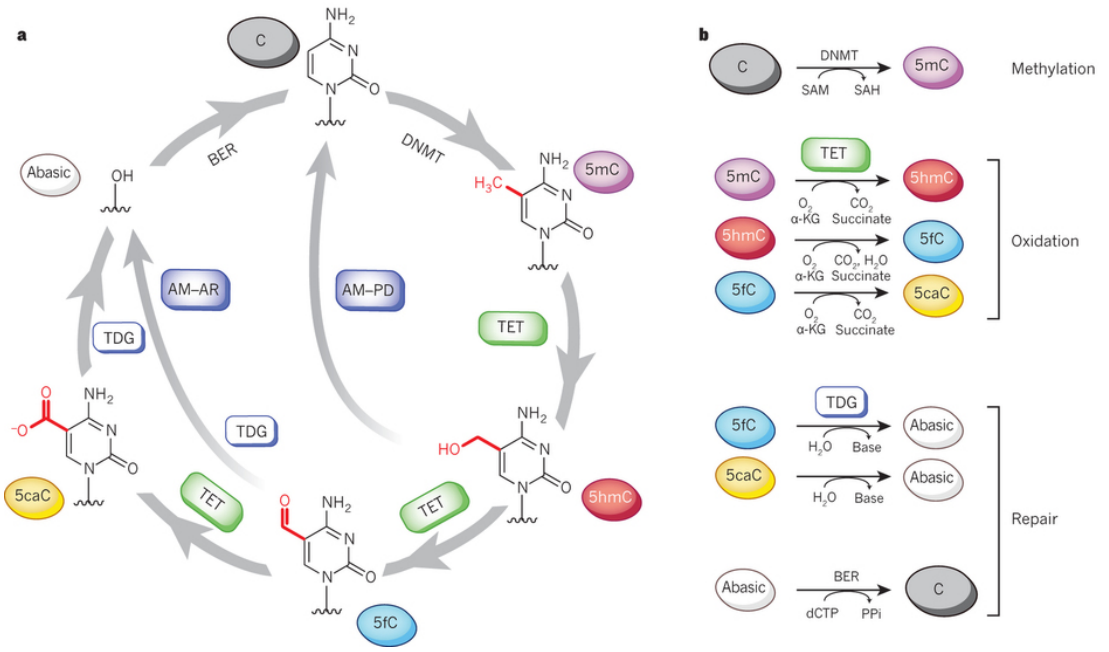


Figure 1.2. Dynamic modifications of cytosine. a. Cytosine (C) converted to 5'methyl-cytosine (5mC) by DNA methyltransferase (DNMT), 5mC oxidized iteratively to 5-hydroxymethylcytosine (5hmC), 5-formylcytosine (5fC), and to 5-carboxylcytosine (5caC); Active modification (AM) – passive dilution (PD): 5hmC is diluted to generate unmodified C; AM – active restoration (AR): 5fC or 5caC is excised by TDG generating an abasic site as part of the base excision repair (BER) process that regenerates unmodified C. **b.** The reactions are shown individually. Figure and legend adapted from [66].

DNA methylation can lead to gene silencing though several mechanisms: methylated CpG sites can directly prohibit transcription factor binding by blocking the DNA binding sites [68] (Figure 1.3a); Methyl-CpG-binding proteins (MBPs) can use co-repressor molecules to repress transcription [69] (Figure 1.3b), or inhibit elongation [70] (Figure 1.3d); interactions between DNMT and histone modification can condense the chromatin and lead to transcriptional repression (Figure 1.3c) [60].

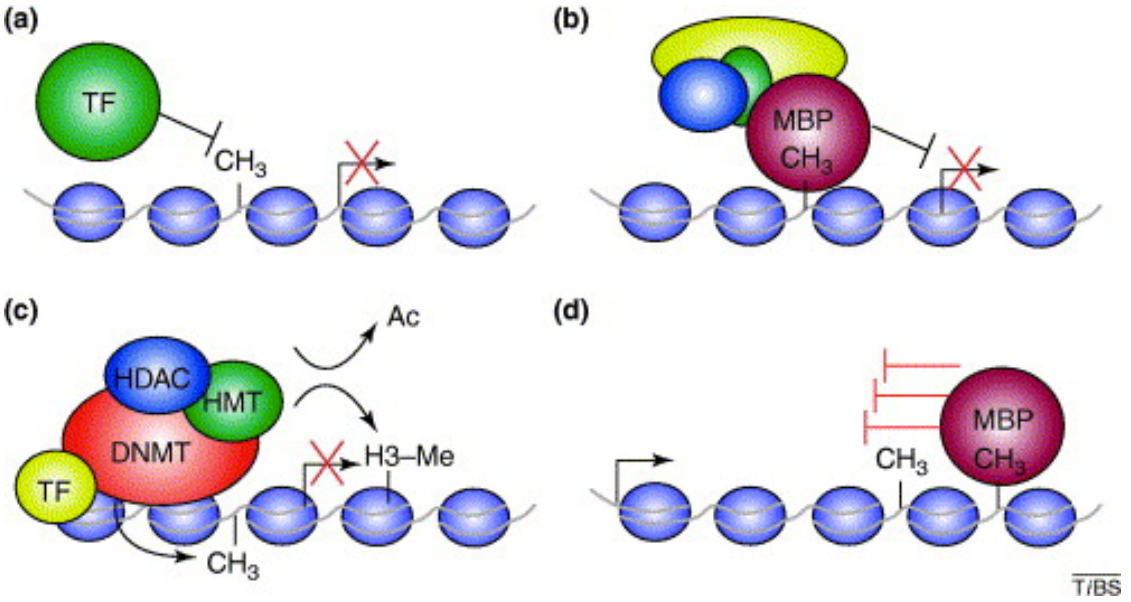


Figure 1.3. Mechanisms of DNA methylation mediated repression. (a) DNA methylation inhibits transcriptional activation by directly blocking the DNA binding sites for transcription factors (TF). **(b)** Methyl-CpG-binding proteins (MBPs) recognize methylated DNA and recruit co-repressor molecules to silence transcription and to modify surrounding chromatin. **(c)** DNMT are linked to histone deacetylase (HDAC) and histone methyltransferase (HMT) and this can lead to chromatin modification and transcriptional repression. **(d)** DNA methylation within the body of genes can effect transcriptional elongation. MBPs might be involved in inhibiting elongation, either directly or by their effects on the surrounding chromatin structure. Figure and legend adapted from [60].

DNA methylation changes are involved in human disease as well as during normal development. Widely studied examples of processes controlled by DNA methylation are X chromosome inactivation [71], genomic imprinting [72], and genome integrity [73].

1.4.3 DNA methylation and disease

Cancer is among the well-known DNA methylation-associated diseases where CpG islands are found to be hypermethylated and general genomic regions are hypomethylated [74,75]. The tumor-suppressor genes are inactivated through methylation of their promoter regions [76]. Hypermethylation

and silencing of genes involved in cell-cycle regulation, chromatin remodelling, cell signalling, transcription and apoptosis favours tumour cells by increasing their genetic instability. Other pathological consequences of deregulation of DNA methylation are characterised in genomic imprinting disorders and repeat-instability diseases [77].

Recently, the role of epigenetics in host-pathogen interactions received attention because of its potentially profound effects on gene expression and its heritable nature [78,79]. Pathogens hijack host cells and control intracellular processes to promote their own survival and evolve phenotypic traits to adapt to the host environment. Conversely host cells induce signals activating the immune response [4]. An effective strategy used by bacteria to alter the host cell is to interfere with the key cellular processes at an early stage. The information flow that gives rise to these processes start in DNA. Hence DNA is an obvious target to reprogram the host cell beneficial for infection. Thus bacteria can affect host cell functions by directly influencing epigenetic factors acting on DNA such as DNA methylation.

In the past three years, there have been a few reported studies of genome-wide changes in the methylation state of host cells. One study examined DNA methylation changes in human dendritic cells infected with a live virulent strain of *Mycobacterium tuberculosis*, the causative agent of tuberculosis in humans [80]. This study showed that host methylation changes were associated with activation of key immune factors. A similar observation was made by Marr *et al.* [81], where the methylation changes in host macrophages infected with the intracellular protozoan *Leishmania* lead to down-regulation of innate immunity, favouring pathogen survival.

There are also reports of *H. pylori*-associated DNA hypermethylation [82–85]. *H. pylori* is a risk factor for gastric cancer [82]. *H. pyroli* is reported to mediate methylation of the E-cadherin (CDH1) tumor suppressor gene promoter [83]. Temporal changes in DNA methylation have been reported in key inflammatory response genes such as CXCL2, IL-1β, NOS2, and TNF-α. This evidence implies that methylation events may contribute significantly to establishing inflammatory responses. In another study it has been proposed that intestinal flow influences the methylation state of the IL-4 gene in intestinal epithelial cells and this results in the down regulation of TLR4 expression [86].

Little is known about bacterial control of host cell methylation beyond these few studies. Yet advances in epigenetic technologies have also made genome-wide profiling of DNA methylation possible. Thus, it is now possible efficiently to identify the repertoire of genes that are differentially methylated as a consequence of bacterial infection. This would not only allow the identification of the role of epigenetic changes in regulating host cell transcription; it could also allow the deciphering of new mechanisms by which bacteria establish disease. Epigenetic changes underlying the host-pathogen relationship could be utilized as biomarkers for assessing the development of diseases. Within the heritable characteristics of these markers, it could also help our understanding of microbial persistence and the evolution of host immunity. By providing new insights into host-pathogen interactions, epigenetic studies can pave the way for novel strategies to prevent or treat disease.

1.5 Trehalose metabolism and virulence

In addition to direct damage to the host caused by virulence factors, a fundamental aspect of pathogenesis is the metabolic regulation as the pathogens adapt to changing conditions (e.g. level of free iron, pH, carbon source and oxygen availability, etc.) in the host [87]. The nature of this complex host-pathogen interaction suggests that metabolic systems are as important as the virulence genes that control membrane receptors and lipids [88]. Pathogenic bacteria can acquire additional metabolic pathways or lose non-essential metabolic functions while undergoing metabolic adaptation to survive within the host [89]. Resistance to antibiotics can also be a result of this metabolic reprogramming [90]. Hence, studying bacterial metabolism is essential for understanding host-parasite interactions and might identify novel targets for anti-microbial therapies.

A relation between trehalose metabolism and virulence traits is being elucidated in pathogenic species [91]. The underlying mechanisms of this association, however, are less clear. While biological experiments often provide a verbal reasoning framework within which to build theories, the complex nature of biological interactions increasingly requires the aid of computational models to comprehensively explain the biological phenomena. To this aim, Chapter 4 first reviews the current knowledge of trehalose metabolism, which then is formalized into computational models and subsequently utilized to make predictions on the physiological behaviour

Trehalose is a sugar present in many organisms such as bacteria, yeast, fungi, plants and invertebrates. Long-established functions of trehalose are energy storage and protection against a variety of stress conditions such as

heat, cold, dehydration and oxidation [92]. Moreover, in mycobacteria and corynebacteria, trehalose is a structural component of cell wall glycolipids. In recent decades, a remarkable variety of functions of trehalose have been shown, including the control of metabolic pathways and growth [93,94].

More recently, the role of trehalose in pathogenic host interactions and virulence has begun attracting widespread interest [91,95–98]. Trehalose accumulation has been found to protect cells against reactive oxygen species [99], which are often used by host cells to kill pathogens. The abolishment of trehalose production results in reduced *in vivo* survival in pathogenic fungi [100–104] and plant pathogens [105]. Most of the studies relating trehalose metabolism and bacterial virulence focus on the *Mycobacterium tuberculosis*, where mutants in trehalose biosynthesis pathways exhibited growth defects and reduced ability to develop chronic infections [95,97].

A number of metabolic pathways are present for biosynthesis and degradation of trehalose in different organisms [106]. Bacteria can synthesise trehalose via five distinct pathways (Figure 1.5.1), two of which are present in eukaryotic organisms [106]. Humans cannot synthesize trehalose; however, humans have the enzyme trehalase required for degradation of trehalose into glucose [107]. Because the trehalose biosynthesis pathways are absent in mammalian cells [108], trehalose related targets have been proposed for novel drug discovery [109].

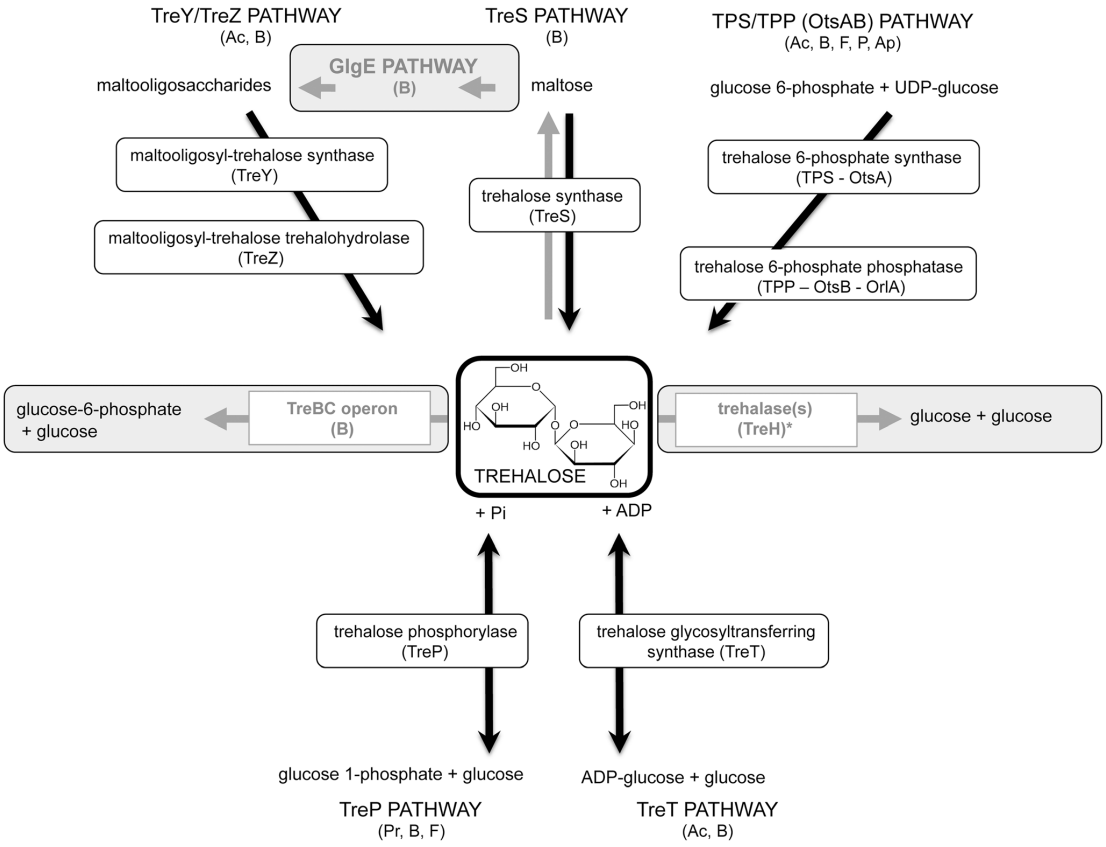


Figure 1.5.1. Metabolic pathways of trehalose synthesis and degradation (highlighted in black and in grey respectively). Ac, Archea; B, Bacteria; F, Fungi; P, Plants; Ap, Arthropods; Pr, Protists. * Trehalase is present in all kingdoms, including mammals. Figure and legend adapted from [91].

1.6 Melioidosis and *Burkholderia pseduomallei*

Burkholderia pseudomallei, a member of the *Burkholderia* genus, is an aerobic, Gram-negative bacterium and the causative agent of the severe disease, melioidosis [110]. Melioidosis occurs after skin inoculation or inhalation of *B. pseudomallei*, which resides in soil, stagnant water, and rice paddies in endemic areas, mainly in Southeast Asia and Northern Australia [111]. Recent reports document that the prevalence of melioidosis is substantially underestimated, with significant implications for public health in many tropical developing countries as well as developed countries [112]. Furthermore,

Centers for Disease Control and Prevention classified *B. pseudomallei* as a category B biothreat agent [113].

B. pseudomallei can infect the skin, spleen, liver, lungs, or can enter the bloodstream spreading throughout the body [111]. Signs and symptoms of melioidosis include cough, chest pain, high fever, headache, and skin abscesses [113]. The inflammatory condition in the lungs, pneumonia, is the common clinical manifestation of melioidosis [114]. As other infections such as tuberculosis commonly affect the lungs, melioidosis is often difficult to distinguish and has been referred to as “the great mimicker” [110]. Melioidosis can occur in acute and chronic forms [115]. Some infected individuals may not develop any symptoms; *B. pseudomallei* can persist in the body and reactivate later in life causing disease [115]. Immunocompromised individuals, e.g. patients with diabetes mellitus or chronic renal failure, are more susceptible [116]. The majority of melioidosis patients develop sepsis [117], a strong immune response with widespread inflammation throughout the body. A variety of currently available antibiotics are ineffective in treating melioidosis and therefore melioidosis has a high mortality rate [113].

B. pseudomallei is capable of surviving harsh environmental conditions such as prolonged nutrient deficiency [118], acidic environments [119], temperature extremes and dehydration [120]. The *B. pseudomallei* genome is composed of two chromosomes: a large chromosome of 4.07 megabase pairs is associated with core functions such as central metabolism and growth and a small chromosome of 3.17 megabase pairs encodes for accessory functions, which may provide the bacterium with the ability to adapt and survive within the host [121].

It has previously been shown that host genes involved in immune response, stress response and cell cycle regulation are differentially regulated upon *B. pseudomallei* infection [122]. Upon infection with *B. pseudomallei*, increased expression of cell surface molecules are observed in human macrophages [123]. Whole-genome transcriptional profiles showing the differentially expressed genes on *B. pseudomallei* infection, including genes related to inflammation, are present in the literature [124].

CHAPTER 2. HOST DNA METHYLATION CHANGES DURING INFECTION WITH *BURKHOLDERIA PSEUDOMALLEI*: IN VITRO INFECTION MODEL

Chapter 2.1 Introduction

At present, very little attention has been given to host DNA methylation changes following bacterial infection. In particular, the time course of such methylation changes during infection has not been addressed.

The research work described in this thesis constitutes the first investigation of the role of host DNA methylation during *B. pseudomallei* infection. In particular, the following research questions represent the main objectives of this work:

(1) To explore whether there are any changes in the host methylome following infection with *B. pseudomallei*.

(2) In cases of proven DNA methylation changes during *B. pseudomallei* infection, to measure the dynamics of methylation changes.

(3) To understand the functional significance of the methylation changes.

(4) To relate methylation changes to gene expression.

An appropriate infection model was essential in capturing the potential role of DNA methylation. In the current research environment, the use of human cells for infection assays with *B. pseudomallei* represents a valid approach [125–129]. The human macrophage cell line U937 was therefore chosen for this work. U937 cells are well established and documented in the literature to study *B. pseudomallei* infection [123,126]. The protocol reported by Lehmann *et al.* [130] was adopted for the setup of the infection model. U937 cells differentiate into macrophages; macrophages provide a good model as they are usually

recruited during infection in response to pathogens to initiate defence mechanisms. Working with a human cell line also enabled the use of microarrays for downstream DNA methylation analysis.

This study was set out as follows:

- *B. pseudomallei* colonies recovered from the differentiated U937 cell monolayers were examined in order to understand the dynamics of *B. pseudomallei* infection in U937 cells. This guided the time points used for the discovery experiment.

- The discovery experiment (referred to as experiment 1) aimed to detect changes in the methylation state of the infected cells compared to the uninfected controls.

- Subsequently, a replication experiment (referred to as experiment 2) was performed. This experiment followed the same protocol of experiment 1 with additional sampling times, allowing for investigation of the time course of the infection.

2.2 Methods

2.2.1 Cell line and infection model

The human leukemic monocyte lymphoma cell line (U937, ATCC CRL-1593.2) was maintained in RPMI 1640 supplemented with 10% foetal bovine serum (FBS) at 37°C. U937 cells were differentiated to macrophage-like cells following exposure to 20 ng/ml (final concentration) of phorbol 12-myristate 13-acetate (PMA) for 48h at 37°C and differentiation evidenced by increased adherence to tissue culture flasks.

Overnight cultures of *B. pseudomallei* K96243 were diluted in L-15 medium and added to differentiated U937 cells at a multiplicity of infection

(MOI) of 10. The dynamics of *B. pseudomallei* infection were established in which *B. pseudomallei* colonies were recovered from the differentiated U937 cell monolayers. At 1h (T1), 2h (T2), 4h (T4) and 24h (T24) post infection, the cells were washed 3 times in warm phosphate buffered saline (PBS) and lysed with 0.1% (vol/vol) triton X-100. Serial dilutions of the cell lysate were plated onto LB agar to determine the intracellular bacterial cell counts. Bacterial loads were measured as colony forming units (CFU) at 2h, 3h, 4h, 5h, 6h, and 24h post infection.

Additionally at T2, the uptake of *B. pseudomallei* K96243 expressing red fluorescent protein (RFP) [131] by U937 cells was enumerated. Cells were washed 3 times with PBS and overlaid with 200µl paraformaldehyde (PFA) 0.4%, ensuring any coverslips were fully immersed. Cells were than incubated at room temperature for 30 minutes. PFA was removed and coverslips were washed twice with PBS for 1 hour for each wash. Coverslips were removed and stained with 4',6-diamidino-2-phenylindole (DAPI). DAPI passes through the cell membrane and binds to DNA. DAPI's blue emission labelled the nuclei of the host cells. Cells were visualized with epifluorescence. Eight fields of view were visualised and the number of uninfected or infected U937 cells were counted. The mean number of cells infected with bacteria was calculated.

The data from this experiment, presented in Section 2.3.1, informed the time points taken for the main experiment in which samples are prepared to measure the DNA methylation.

A total of eight tissue culture flasks (four for infected cells and four for uninfected controls) were prepared for the first set of samples for the DNA methylation study. An overview of the experimental setup is illustrated in Figure 2.2.1. Uninfected controls were overlaid with L15 medium only. The cells were

then incubated at 37°C for 2h to allow infection. The cells were washed 3 times with PBS and incubated with fresh L15 medium containing 1mg/ml kanamycin for 2h to kill extracellular bacteria. After 2h, the macrophage cells were held in fresh media containing 250µg/ml kanamycin to supress the growth of extracellular bacteria. At T2 and T4, the cells were washed 3 times in warm PBS and lysed with 0.1% (vol/vol) Triton X-100. DNA was isolated using an AllPrep kit (Qiagen) and stored at -80ºC until required. DNA yield was measured using a Nanodrop instrument with measurements between 22.8–50.6 ng/ul. This experiment was repeated, ultimately resulting in two technical replicate and two experimental replicate DNA samples for each time point, T2 and T4 (a total of 16 samples). Although the Qiagen AllPrep method allows simultaneous isolation of DNA and RNA, RNA was not collected in this experiment. This was due to difficulties in handling high number of flasks in Biosafety Level 3 Laboratory. The challenges that emerged from this experiment were taken into consideration in designing the second experiment.

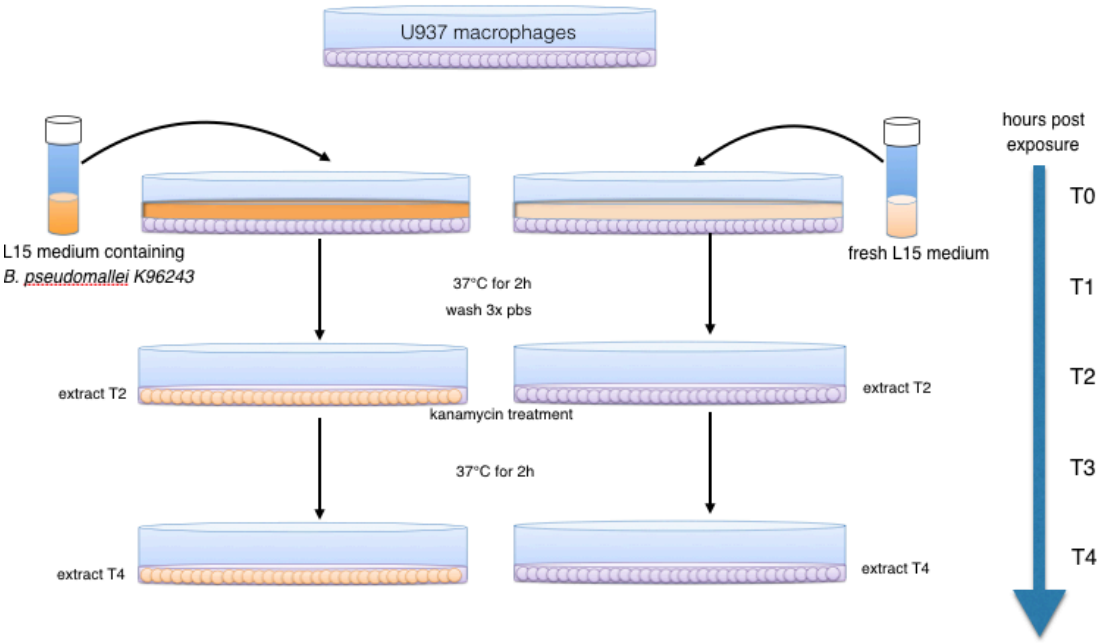


Figure 2.2.1. Experimental setup for macrophage infection.

A second experiment was designed to enable verification of the results of the first experiment and to allow additional sampling times and collection of RNA. Similar procedures were followed as described above and DNA and RNA was collected from uninfected and infected cells at 1h (T1), 2h (T2), 3h (T3) and 4h (T4) post infection. At T3 and T4 two types of controls were collected; uninfected cells administered with kanamycin and uninfected cells without kanamycin. For every group, 1 sample was obtained (a total of 10 DNA samples and a total of 10 RNA samples).

2.2.2 DNA methylation profiling

The DNA methylation profile of the host macrophage DNA was determined using the Infinium HumanMethylation450 BeadChip (450K) (Illumina Inc.). The 450K BeadChip interrogates DNA methylation at >480,000 individual CpG sites (loci) on the human genome.

The first step is bisulfite treatment. Bisulfite treatment converts unmethylated cytosines to uracil, which is then amplified as thymine during whole genome amplification [132]. Methylated cytosine residues are left unaffected. This allows the sequencing technologies to distinguish between methylated and unmethylated cytosine residues [133].

The bisulfite-converted-amplified DNA products are then applied to the chip and hybridised either to the methylation specific probe or to the unmethylation probe. Single-base extension of the probe incorporates a fluorescently labelled dideoxynucleotide (ddNTP). The chip is then scanned and the fluorescent signal is measured.

HumanMethylation450 BeadChip technology uses two different types of probes (Infinium Type I and II probe designs are illustrated in Figure 2.2.2). The Infinium Type I Probes have two separate probe sequence per CpG site - one for methylated CpG and one for unmethylated CpG. Thus, unmethylated and methylated DNA fragments hybridise to two different beads and produce a red signal. Infinium Type II Probe measures unmethylated and methylated DNA fragments by the same bead but produce a green signal for the methylated fragment and a red signal for the unmethylated fragment. Type II probes are more efficient as they use one probe sequence per CpG site: however Type I probes perform better in measuring CpG dense regions [134].

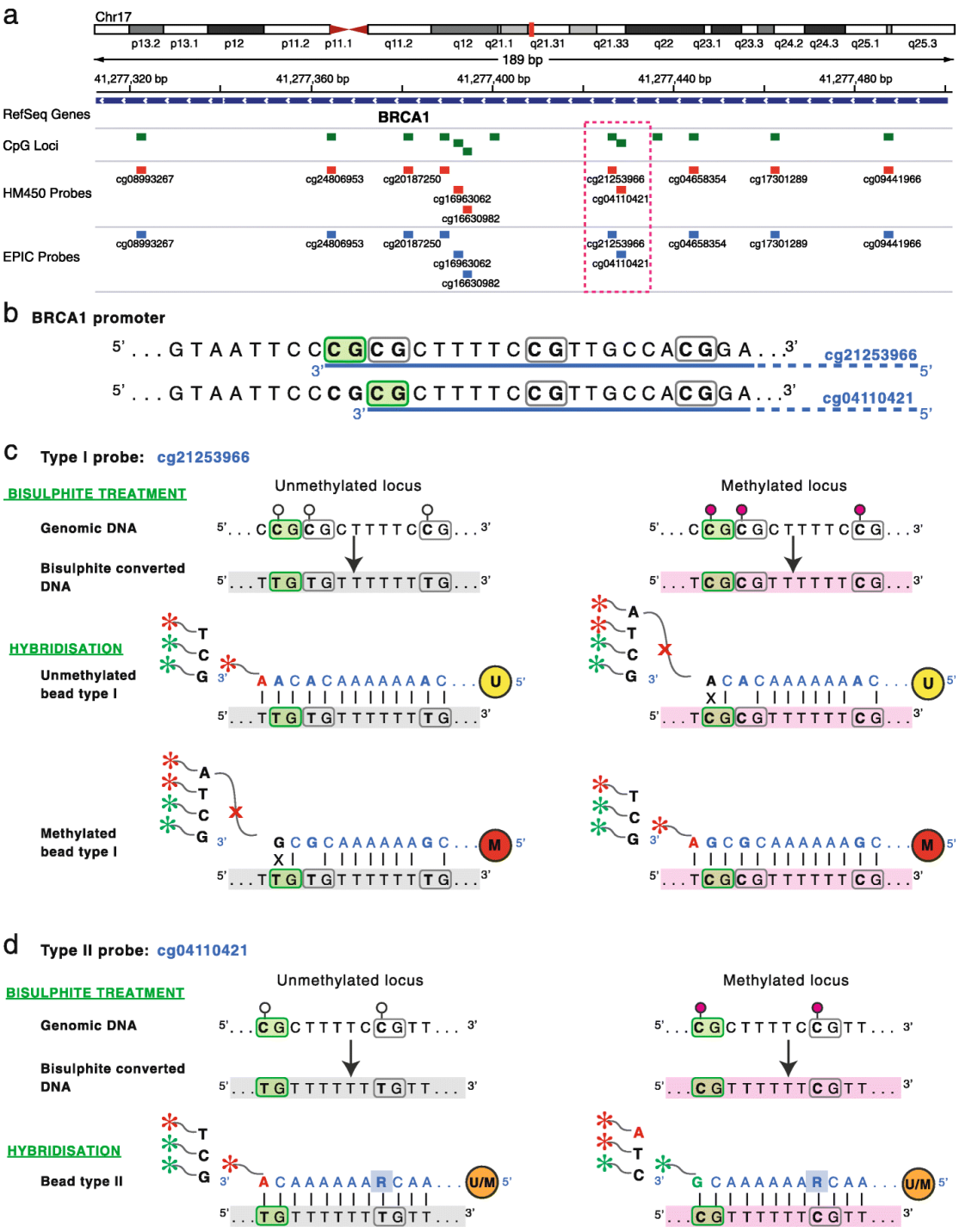


Figure 2.2.2. The design of Infinium methylation assays (See legend on next page)

(See figure on previous page) (a) The genomic location of the gene, *BRCA1*. CpG sites (CpG Loci) are marked in green. The probes present in HumanMethylation450 (HM450) and MethylationEPIC (EPIC) platforms are marked in red and blue, respectively. Two probes targeting adjacent CpG sites in the *BRCA1* promoter were circled in pink dotted line. These are two different types of probes (Infinium Type I and II probes). (b) Infinium I (cg21253966) and Infinium II (cg04110421) probes; the targeted CpG sites are highlighted in green. Each probe is designed to hybridise a 50 bp DNA sequence, underlined in blue, downstream of the targeted CpG site. (c) In bisulfite conversion the unmethylated cytosine (C) is converted into uracil and later amplified as thymine (T). The unmethylated signal detection for the cg21253966 probe is schematically represented on the left panel. The unmethylated bead probe (U) sequence is designed to match bisulfite converted DNA sequence of the unmethylated locus. The hybridisation of a bisulphite converted unmethylated DNA fragment to the bead enables single base extension and incorporation of a ddNTP labelled nucleotide matching the nucleotide immediately upstream of the target CpG site; in this case incorporation of an A nucleotide and signal detection in the RED channel. Hybridisation of the methylated bead probe (M), on the other hand, results in mismatch at the 3′ end of the probe and inhibition of single base extension. Detection of the methylated signal, shown on the right panel, follows similar steps. (d) For Infinium II probes, the unmethylated and methylated signals are measured by the same bead (U/M). The bead probe sequence is designed to match bisulphite converted DNA of both the methylated and unmethylated locus. This is achieved by making the cytosine of the target CpG site the single base extension locus and replacing cytosines of all other CpG sites within the probe sequence with degenerate R bases that hybridise to both T (representing unmethylated and converted cytosine) and C (representing methylated and protected cytosine) bases. Single base extension and incorporation of ddNTP labelled A nucleotide matching the unmethylated and converted cytosine at the target CpG site results in signal detection on the RED channel. The detection of the methylation signal, shown on the right panel, is the same except that in this case single base extension results in incorporation of ddNTP labelled G nucleotide matching the methylated and protected cytosine at the target CpG site and signal detection on the GREEN channel. Figure and legend adapted from Pidsley *et al.* [134].

To profile the samples on the array, 500ng of genomic DNA was sodium bisulfite converted using the EZ-96 DNA Methylation kit (Zymo research, CA, USA). Replicates were processed together on the same array to reduce batch effects. The data were extracted using the GenomeStudio (2010.3) methylation module (1.8.5).

2.2.3 Data acquisition and normalisation

All quality control checks were performed using R (version 3.2.3). Signal intensities were imported into R using the *methylumi* package. For each sample, the distribution of the methylated and unmethylated signals were plotted to check whether there were any extreme outliers.

For each CpG site a β-value was calculated, which is the ratio of the methylated probe intensity (M) to the overall intensity (sum of methylated (M) and unmethylated (U) probe intensities) [135]. For an i^{th} interrogated CpG site:

$$\beta_i = \frac{\max(M_i, 0)}{\max(M_i, 0) + \max(U_i, 0) + \alpha}$$

where M_i and U_i are the intensities measured for the i^{th} methylated and unmethylated probes, respectively. For Infinium Type I Probe, the M and U intensities are signals produced by two different bead types and are reported in the same colour (red). For Infinium Type II Probes, M and U are the intensities of the green and red fluorescent signals, respectively. Any negative values that might arise from global background intensity subtraction are reset to 0. When both methylated and unmethylated probe intensities are low, a constant offset α (α = 100 as recommended by Illumina [135]) was added to the denominator to regularise the β-value. Ultimately, the β-value between 0 and 1 reflects the methylation level of each CpG site: a value of zero represents that all measured loci in the sample are unmethylated and a value of one represents that all measured loci in the sample are methylated.

Quality control checks and quantile normalisation were implemented using R *WateRmelon* package [136]. Samples with more than 1% of sites with a detection *p*-value greater than 0.05 were removed, as were probes with 1% of samples with a detection *p*-value greater than 0.05 (following guidelines

established by Bibikova *et al.* [137]). Probes were removed if they had a bead count less than 3 in 1% of samples. Cross-hybridizing probes were removed [138], leaving 425496 probes for analysis.

2.2.4 Data analysis

Reasoning that larger differences are potentially more biologically meaningful [139], differentially methylated positions (DMPs) between infected and control cells were identified using a Δβ cut-off of 0.1. The analysis pipeline is summarised in Figure 2.2.3.

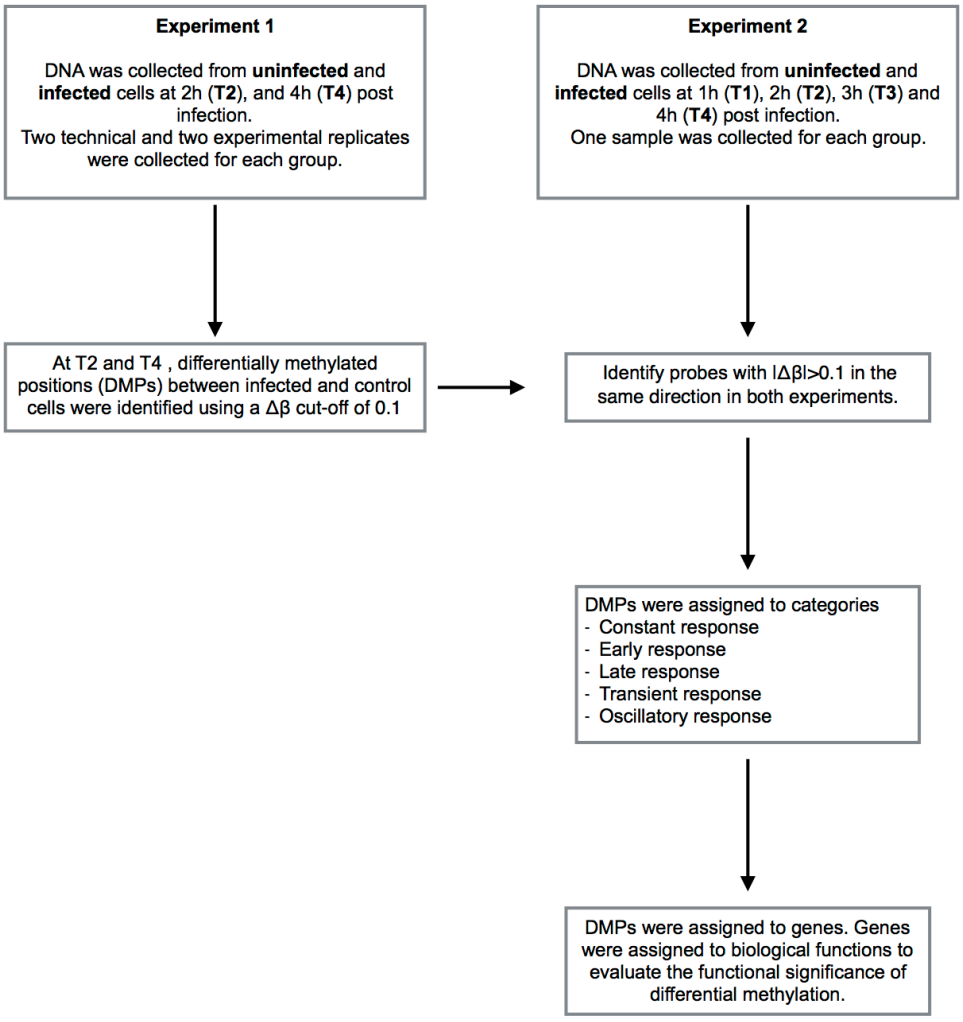


Figure 2.2.3. Schematic of the data analysis pipeline.

DMPs identified at T2 and T4 in the first experiment, were compared to the corresponding probes at the same time point in the second experiment. Probes exhibiting a $\Delta\beta$ change in the same direction with $|\Delta\beta|>0.1$ were taken for further analysis. Correlations of $\Delta\beta$ values between the two experiments were measured using Pearson's correlation coefficient. This set of replicated DMPs was tested for enrichment of genomic regions using a two-tailed Fisher's exact test, comparing DMPs to the frequency of all probes on the Illumina 450K array.

DMPs with consistent up or down-methylation at all time points T1, T2, T3 and T4 were identified. The temporal methylation patterns were assigned to categories based on the change in score at successive time points. As many of the differential methylation patterns found in disease or environmental factors are characterized by smaller changes, in the range of 5% [81], this criterion was selected to identify temporal patterns of differential methylation. Signatures with more than 5% differential methylation at all time points were classified as, constant hyper- (+5%) or constant hypo-methylated (-5%) probes. Signatures with more than 5% differential methylation at early time points (T1, T2, T3) and no change at the T4 time point were classified as early response probes. Signatures with no response at the T1 time point and with more than 5% differential methylation at later time points were classified as late response probes. Signatures with no response at T1 and T4 and more than 5% differential methylation at T2 or T3 were classified as transient responses. The remainder of the patterns were classified as oscillatory probes.

Candidate genes were assigned to the probes using the GREAT software [140], which allots genes to genomic regions taking into account the functional significance of *cis*-regulatory regions. A probe level analysis was

employed using a custom script to determine the functional significance of the differentially methylated positions in experiment 1. Genes were assigned to functions using the Gene Ontology. Gene ontology annotations were downloaded from the Gene Ontology Consortium database [25]. Probes were matched with Gene Ontology terms.

To measure the association between the top differentially methylated probes and a given biological function, the right-tailed Fisher Exact Test [141] was used where the analysis is applied to focus probes against all probes, as outlined in the contingency table (Table 3.2.1). For each function (gene ontology term), a *p*-value is calculated, which is a measure of the likelihood that the association between the set of focus probes (here, the set of differentially methylated probes) and the given function is due to random chance. *p*-values less than 0.05 are considered to be statistically significant, non-random associations. The odds ratio (OR) is another measure of association and can be calculated from Table 3.2.1 as,

$$OR = \frac{n_{11}/n_{21}}{n_{12}/n_{22}}$$

n_{11} is the number of probes that are associated with the given function in the set of focus probes,

n_{21} is the number of probes that are not associated with the given function in the set of focus probes,

n_{12} is the number of all probes on the array that are associated with the given function,

n_{22} is the number of all probes on the array that are not associated with the given function.

The greater the odds ratio is, the more significant is the association.

Table 3.2.1. Contingency table. For each functional term a contingency table is constructed, where the number of probes (n) that are associated and not associated with the function are enumerated for the focus set and the set of all probes.

	Focus probes	All probes	Row totals
Probes associated with a function	n_{11}	n_{12}	row_1
Probes not associated with that function	n_{21}	n_{22}	row_2
Column totals	col_1	col_2	N

An alternative method for gene ontology term enrichment analysis was performed using the Bioconductor package GOseq [142]. GOseq employs a similar method to probe level analysis explained previously. Both methods measure gene ontology enrichment by Fisher's exact test, while the probe level analysis counts the number of probes for a given ontology term, GOseq counts the number of genes. GOseq method is adapted to determine the functional significance of the conserved iDMPs. Enriched gene ontology terms are selected based on the criteria of having a *p*-value<0.05.

Redundant gene ontology terms were removed and the non-redundant gene ontology terms were clustered using semantic similarity measures (the *simRel* score) with the REVIGO software tool [143].

2.2.5 Measuring gene expression using Real-time PCR

100ng total RNA was reverse transcribed from each sample using the Superscript VILO cDNA synthesis kit (Invitrogen, Paisley, UK). Quantitative

real-time PCR assays specific to the DDHA2 and TOLLIP transcripts (Assay identifiers Hs00203889_m1 and Hs01553188_m1 respectively) were purchased from Life Technologies (Paisley, UK). PCR reactions contained 2.5ml TaqMan Universal Mastermix (no AMPerase) (Applied Biosystems, Foster City, USA), 0.9mM each primer, 0.25mM probe and 1ml cDNA in a total volume of 5ml. Samples were tested in three technical replicates. PCR conditions were a single cycle of 95°C for 10 minutes followed by 50 cycles of 95°C for 15 seconds and 60°C for 1 minute. The amount of each transcript was then quantified in RNA samples relative to the endogenous control genes beta microbglobulin (GUSB; assay identifier Hs00984230_m1) and beta glucuronidase (GUSB; Assay identifier Hs00939627_m1). These control genes were chosen on the basis of their stability under test conditions, which was determined empirically. Quantifications were carried out by the Comparative Ct approach and data were normalised to the median value for samples unexposed to bacteria.

2.3 Results

2.3.1 Infection Model

To understand the dynamics of *B. pseudomallei* infection, U937 cells were infected with *B. pseudomallei* K96243 expressing red fluorescent protein (RFP). U937 cells were matured and differentiated to adopt the characteristics of mature human macrophages [123]. The cells were activated with interferon gamma (IFN-γ). The activation of macrophages with IFN-γ dramatically enhances the ability of the cells to control infection with *B. pseudomallei* [144]. At 2 hours (T2) post infection, 74% of the human cells were infected; most infected cells were infected with multiple bacteria (Fig. 2.3.1 - the blue stain shows the nuclei of the host cells and the red stain shows the bacteria). The number of intracellular bacteria was 10^6 CFU/ml at T2 (Fig. 2.3.2). After 4 (T4) or 6 hours (T6) post infection, the number of intracellular bacteria had declined to 10^3 CFU/ml. No intracellular bacteria were recovered at 24 hours post infection (T24). This experiment guided the design of the first experiment to measure DNA methylation. As there were no previous reports on when the DNA methylation changes might be occurring, for our discovery experiment (Experiment 1) two time points were chosen - a first time point when the infection is established (T2) and a later time point when there is reduction in the number of intracellular bacteria (T4). As the number of viable bacteria level off after 4 hours post infection (T4), a similar response would be expected at T4 and after T4.

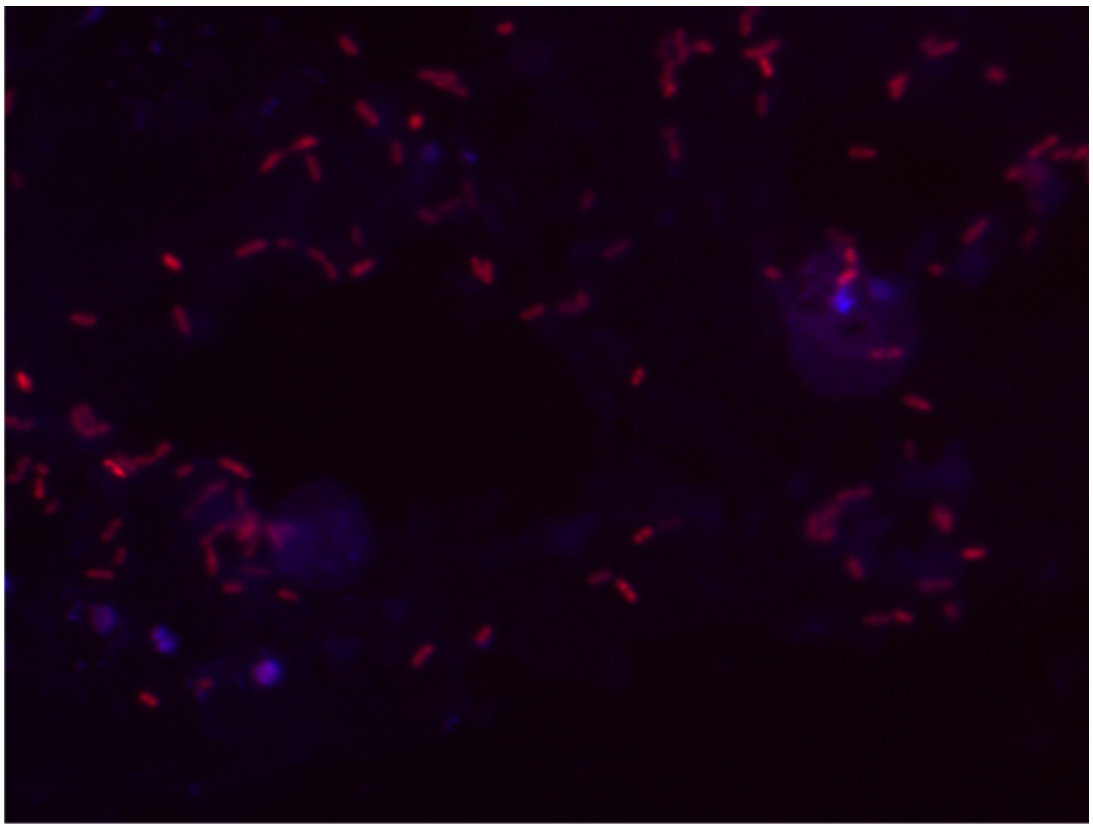


Figure 2.3.1. Infected cells at T2. The proportion (74%) of cells infected at T2 determined by microscopy. Blue stain represents the nuclei of the host cells and the red stain represents the bacteria.

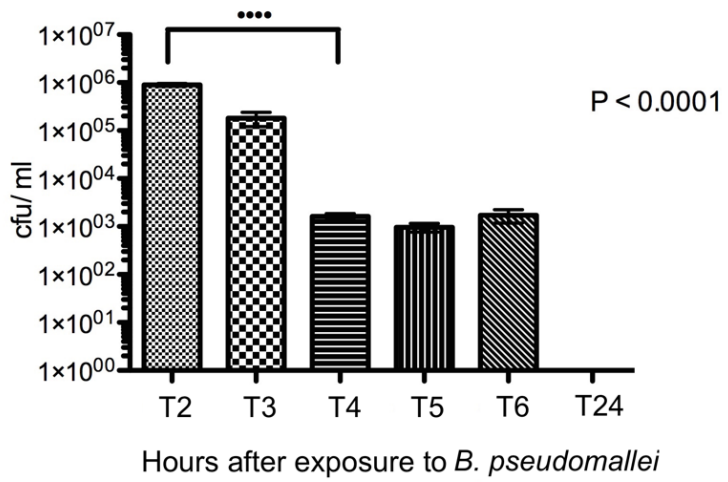


Figure 2.3.2. Bacterial loads at different time points in U937 cells infected with *B. pseudomallei* (MOI = 10). Bacterial load was measured as colony forming units (CFU).

2.3.2 DNA extraction

Simultaneous isolation of DNA and RNA was practiced on test samples using Trizol. The test samples were *B. pseudomallei* infected mouse macrophages (2h post infection) from another study conducted by Olivia Champion. Signs of degradation of the DNA samples and contamination were observed (DNA quantified using Nanodrop, Fig. 2.3.3). The project collaborators found similar problems when using DNA samples that had been stored in Trizol solutions. According to current status of the art, the Qiagen AllPrep method is established as an improved method for preparing genomic DNA for methylation studies compared to Trizol extraction [145]. The use of Qiagen AllPrep overcame the DNA degradation and contamination issues.

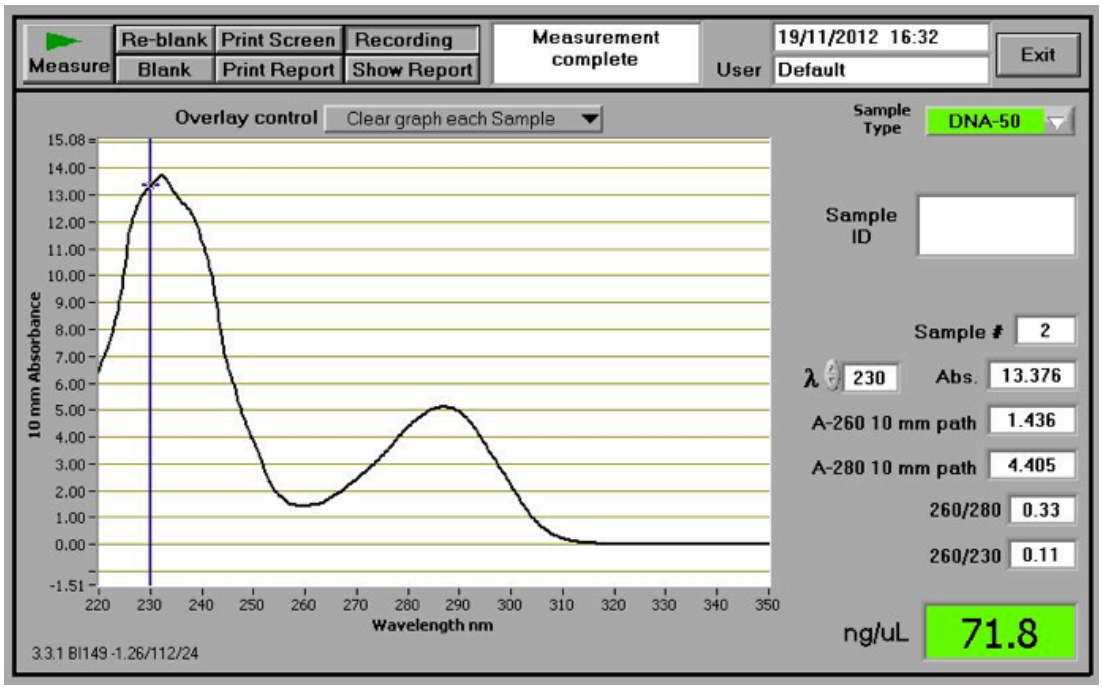


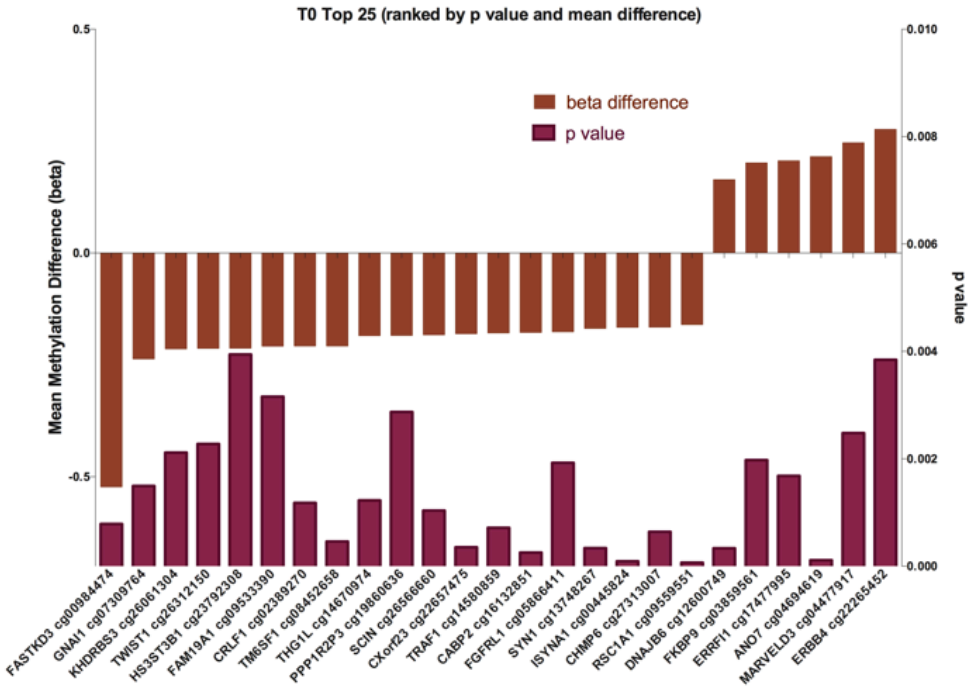
Figure 2.3.3. Quantification of DNA concentration and quality control. DNA isolated using the Trizol extraction method. DNA peak measured by the Nanodrop instrument.

2.3.3 DNA methylation discovery experiment (Experiment 1)

DNA from the infected or uninfected (controls) U937 cells was collected from two technical replicates and two experimental replicates at the selected time points T2 and T4 as explained in Section 2.3.1. Although initially, the study was planned to simultaneously measure DNA methylation and gene expression levels, RNA was not collected in Experiment 1. This was due to experimental difficulties arising from organizing a line of set-ups to simultaneously collect high number of samples in Biosafety Level 3 laboratory. The challenges faced in this experiment were taken into consideration for the design of the follow-up experiment in which RNA samples were collected.

DNA methylation was quantified using an Illumina 450K HumanMethylation array, with pre-processing, normalisation and stringent quality control undertaken as explained in Section 3.2.3. Differentially-methylated positions (DMPs) between infected and uninfected cells at T2 or T4 were identified. For each position (CpG sites), the significance of differential methylation was measured using Student's *t*-test. The positions were ranked by mean methylation difference and *p*-value. Widespread large changes were identified in host cell DNA methylation following infection, top 25 positions were reported in Fig. 2.3.4.

(a)



(b)

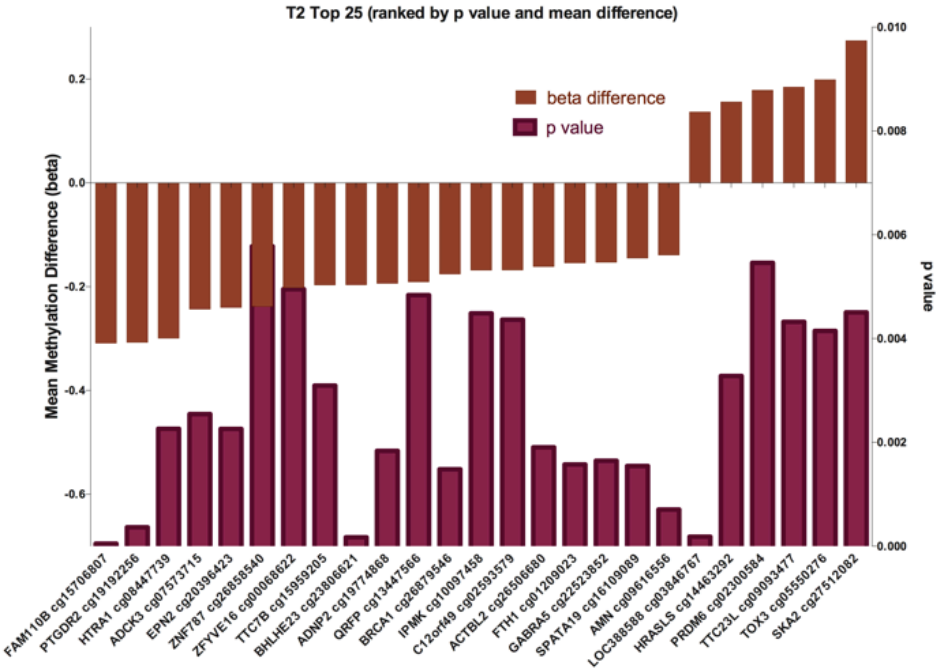


Figure 2.3.4. Top 25 genes at T0 and T2. a) Top 25 genes ranked by mean methylation difference and *p*-value at T2. b) Top 25 genes ranked by mean methylation difference and *p*-value at T4. The x-axis is labelled with the gene name and the CpG cite ID. The left y-axis corresponds to the difference in mean methylation levels between infected and uninfected cells ($\Delta\beta$ = mean $\beta_{infected}$ -

mean $\beta_{uninfected}$). The right y-axis reports the *p*-values measured using Student's *t*-test.

Next, the uninfected samples at T2 were compared to uninfected samples at T4. Ideally one would expect that for control samples, the methylation levels do not change in time, so that the only factor affecting the methylation state is infection. However, there were a few CpG sites with methylation changes in uninfected cells from T2 to T4. These changes could be due to the antibiotic being introduced after the T2 samples had been taken. This would not pose an issue as our main focus is the comparison between infected and uninfected cells. Thus, when the infected and uninfected cells at T4 were compared, both groups contain antibiotic and its effect is cancelled. Nonetheless, uninfected controls with and without kanamycin were included for the replication experiment to explore any effects the antibiotic stress can have on the host DNA.

Different *p*-value cut-offs were applied to enumerate the DMPs in Table 2.3.1. A high number of *t*-tests being performed simultaneously can result in a large number of observations, which will have low *p*-values due to chance. To correct for these random events that falsely appear significant (false positives), multiple testing correction methods are frequently applied [146]. A statistical confidence measure is reported using false discovery rate (FDR) estimation [147], summarized in a *q*-value. The *q*-values were calculated using the *qvalue* package in R [148]. None of the CpG sites had a significant *q*-value (*q*-value < 0.05 considered significant as it is a widely accepted threshold for genomewide studies [149]). However this is not surprising, as CpG sites are not independent; many of the multiple testing algorithms are too stringent, especially considering our small sample size [150]. Based on the statistical or effect size cut-offs

applied, a different interpretation can be given to the data [151]. Within the limitations of this study (i.e small sample size), in this chapter DMPs were identified on the basis of reproducible large methylation differences to maximize the chances of obtaining biologically meaningful results. In Chapter 3, alternative approaches are employed such as regional analysis, in which differential methylation is considered significant when identified on large genomic regions rather than on single CpGs.

Table 2.3.1. Number of DMPs identified for different *p*-value cut-offs. *p*-values were calculated using Student's *t*-test.

p-value	<0.0001	<0.001	<0.01	<0.025	<0.05	<0.1	<1
Number of probes at T2	27	274	2853	7649	15989	34713	456107
Number of probes at T4	18	133	1635	4694	10670	25139	456107

Mean methylation differences at T4 were plotted against the differences at T2 in Figure 2.3.5. Probes with conserved mean methylation differences greater than 0.1 in the same direction at both time points T2 and T4 (coloured in red in Figure 2.3.5) were selected as the focus probes. GREAT gene annotation was used to annotate all probes with ontology terms downloaded from Gene Ontology. To measure the association between the differentially methylated probes and a given biological function, the right-tailed Fisher's Exact Test was used. *p*-values less than 0.05 were considered statistically significant. Top 25 enriched functions were listed in Table 2.3.2.

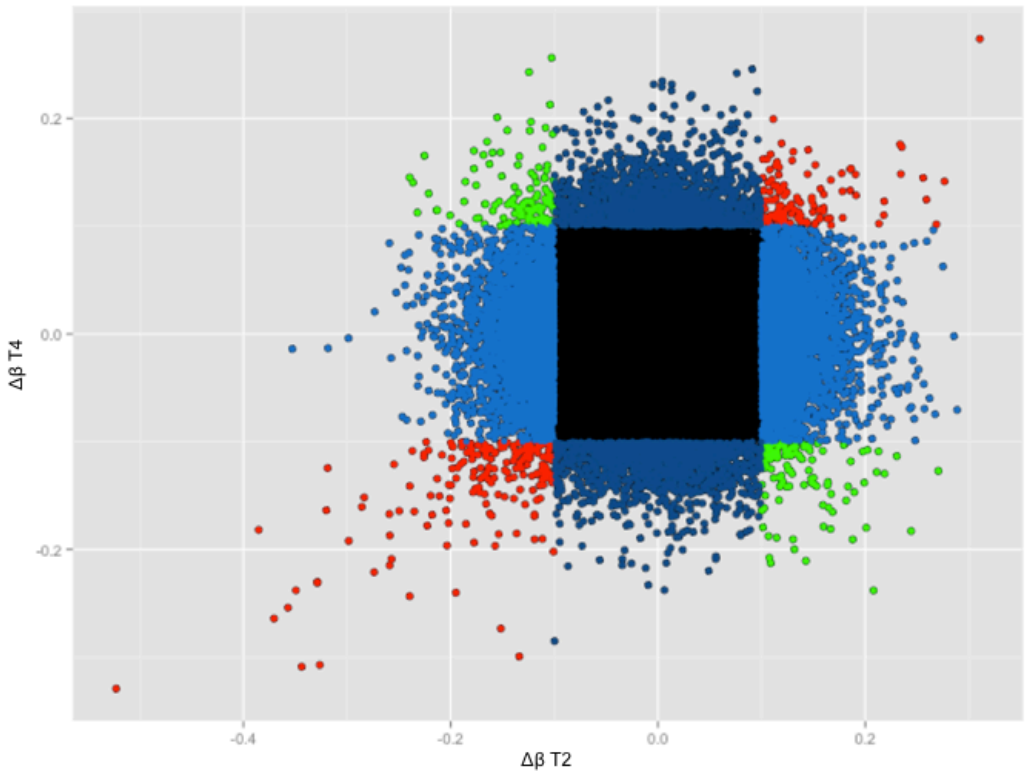


Figure 2.3.5. Scatter plot of the mean methylation (beta) difference (Δβ T2 vs Δβ T4). |Δβ| values greater than 0.1 are highlighted. Red - positive correlation (|Δβ|>0.1 at both time points in the same direction); green - negative correlation (|Δβ|>0.1 at both time points in the opposite direction); blue – |Δβ|>0.1 at one time point but not the other.

Table 2.3.2. Functional annotation of the top probes conserved in the cells colture experiment at time points T2 and T4.

Gene Ontology ID	Number of all probes associated with the given GO	Number of probes in the focus set associated with the given GO	OR	*p*-value	GO_name
GO:0070888	731	5	10.05917789	0.000181624	E-box binding
GO:0005509	13087	21	2.456635913	0.000356606	calcium ion binding
GO:2001022	1395	6	6.340970608	0.000487797	positive regulation of response to DNA damage stimulus
GO:1901723	217	3	20.1611615	0.000518064	negative regulation of cell proliferation involved in kidney development
GO:0050811	230	3	19.01791558	0.000611838	GABA receptor binding
GO:0034599	2678	8	4.429600008	0.000649362	cellular response to oxidative stress
GO:0071837	638	4	9.17284358	0.001124027	HMG box domain binding
GO:0032878	286	3	15.29226573	0.001137529	regulation of establishment or maintenance of cell polarity
GO:2000352	290	3	15.08127142	0.001183177	negative regulation of endothelial cell apoptotic process
GO:0001806	70	2	41.48072019	0.001189611	type IV hypersensitivity
GO:0002524	70	2	41.48072019	0.001189611	hypersensitivity
GO:2000606	70	2	41.48072019	0.001189611	regulation of cell proliferation involved in mesonephros development
GO:2000607	70	2	41.48072019	0.001189611	negative regulation of cell proliferation involved in mesonephros development
GO:2000702	70	2	41.48072019	0.001189611	regulation of fibroblast growth factor receptor signaling pathway involved in ureteric bud formation
GO:2000703	70	2	41.48072019	0.001189611	negative regulation of fibroblast growth factor receptor signaling pathway involved in ureteric bud formation
GO:2000733	70	2	41.48072019	0.001189611	regulation of glial cell-derived neurotrophic factor receptor signaling pathway involved in ureteric bud formation
GO:2000734	70	2	41.48072019	0.001189611	negative regulation of glial cell-derived neurotrophic factor receptor signaling pathway involved in ureteric bud formation
GO:0043204	1155	5	6.355433096	0.001401425	perikaryon
GO:1901722	323	3	13.54281884	0.001604214	regulation of cell proliferation involved in kidney development
GO:0002572	86	2	33.72446425	0.001768647	pro-T cell differentiation
GO:0000266	341	3	12.82075795	0.001868625	mitochondrial fission
GO:0031536	91	2	31.88605423	0.001972052	positive regulation of exit from mitosis
GO:0005134	93	2	31.22282624	0.002056376	interleukin-2 receptor binding
GO:0035898	94	2	30.88405484	0.00209917	parathyroid hormone secretion
GO:0060995	94	2	30.88405484	0.00209917	cell-cell signaling involved in kidney development

2.3.4 Validation with pyrosequencing

A technical verification for the 450K Human Methylation Array was designed to target the two top methylated probes detected in Experiment 1. The top methylated probes for time point T2 and T4 were annotated with genes *FASTKD3* and *FAM110B* respectively. Pyrosequencing assays were designed using the PyroMark Assay design software (Qiagen). The initial steps of DNA methylation analysis by pyrosequencing are bisulfite conversion and PCR. To optimize the PCR, bisulfite conversion and PCR were performed on test samples. Figure 2.3.6 shows the PCR products amplified using the two primers (*FASTKD3* and *FAM110B*). Using the optimised PCR conditions, PCR amplification was performed successfully on bisulfite treated DNA samples prepared in Experiment 1 as shown in Figure 2.3.7. Pyrosequencer monitors the real-time incorporation of nucleotides and signals release of pyrophosphate molecules during DNA elongation, visualises peaks corresponding to the nucleotide sequence. Pyrosequencing analyses for these samples were not successful as peak detection failed when the PCR products were run on the pyrosequencer.

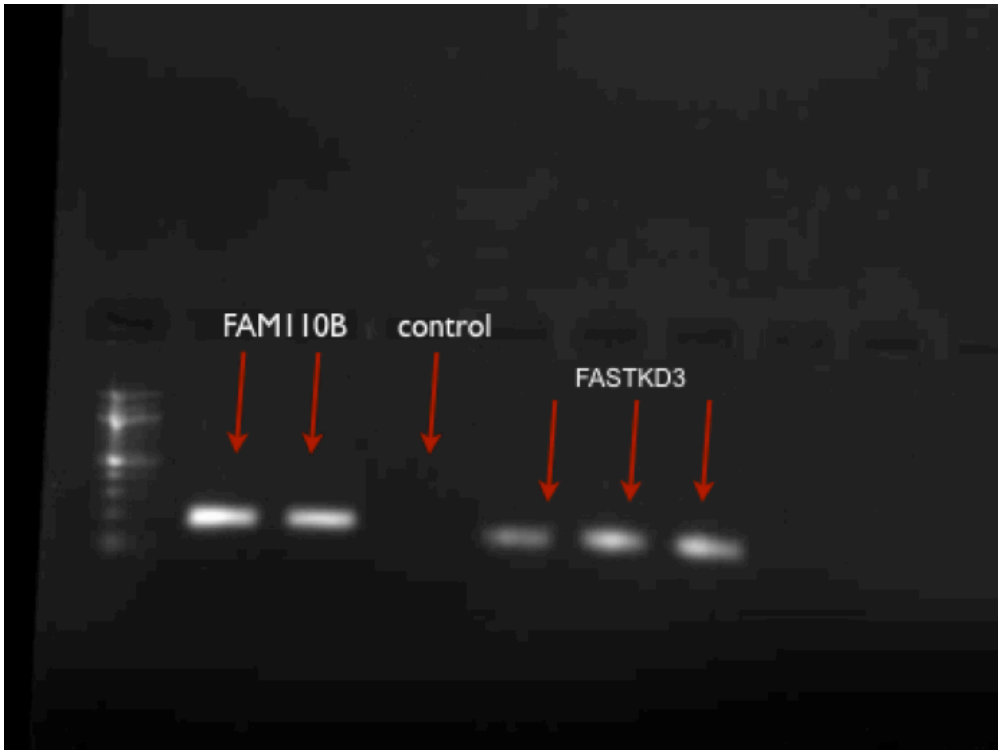


Figure 2.3.6. Test samples - PCR products amplified using the primers indicated.

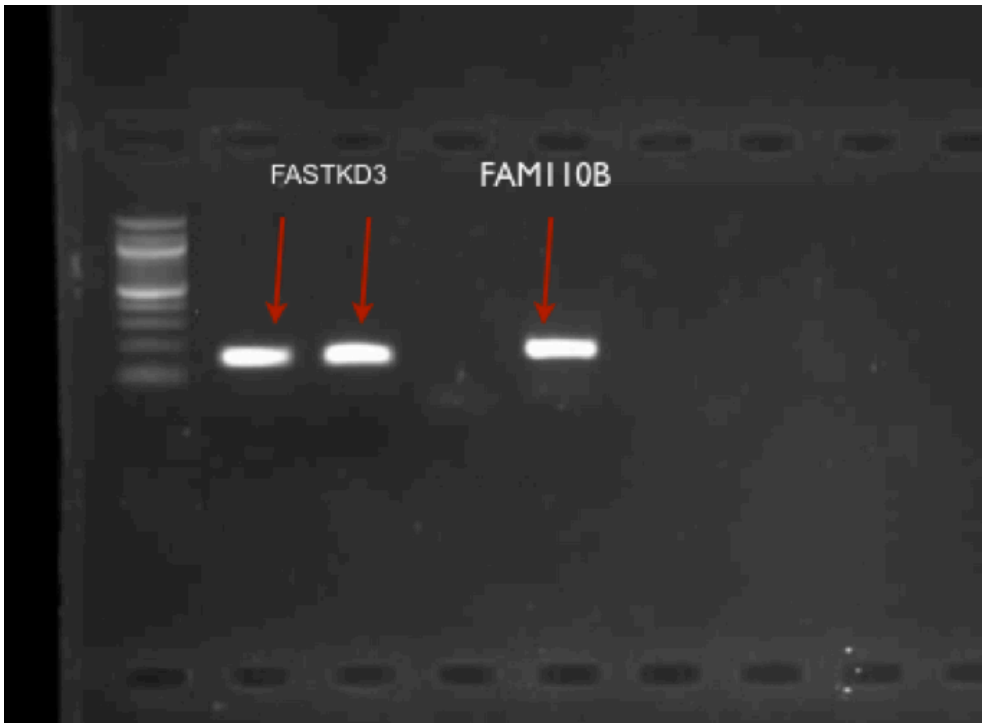


Figure 2.3.7. U937 macrophage samples - PCR products amplified using the primers indicated.

A second line of investigation was employed for pyrosequencing analysis. *B. pseudomallei* infected and control mice lung tissue samples were donated by a collaborator who had already generated transcriptomic data [152]. These were from mice acutely infected with *B. pseudomallei* by the intranasal route day 1 and day 3 post infection. Uninfected control mice received pyrogen-free saline (PFS). DNA was isolated from a total of 16 tissue samples using Trizol extraction and were quantified using Nanodrop (Table 2.3.3).

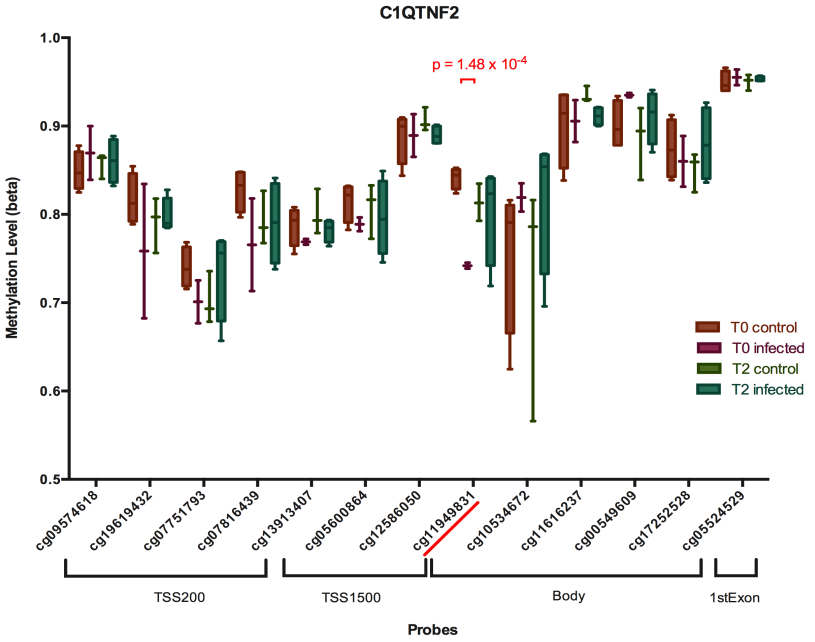
Table 2.3.3. DNA from mouse lung samples.

Experiment	Harvest (day post infection)	Mouse Lung samples	DNA Conc. ng/ul
LC1111	1	L1 PFS	47.8
LC1111	1	L2 PFS	29.4
LC1111	1	L3 PFS	27.5
LC1111	1	L4 PFS	31.5
LC1111	1	L15 Bps	53.2
LC1111	1	L16 Bps	44.6
LC1111	1	L17 Bps	11.4
LC1111	1	L18 Bps	26.6
LC1111	3	L25 PFS	73.5
LC1111	3	L26 PFS	48.4
LC1111	3	L27 PFS	50.7
LC1111	3	L28 PFS	89.4
LC1111	3	L39 Bps	87.4
LC1111	3	L40 Bps	72.7
LC1111	3	L41 Bps	195.4
LC1111	3	L42 Bps	121.8

Two genes were selected as targets for the pyrosequencing analysis, C1QTNF2 and TRIM10. These genes exhibited differential methylation in Experiment 1 in this study (methylation measurements shown in Figure 2.3.8) and also were differentially expressed on the mouse transcriptomic data. The pyrosequencing assay for TRIM10 failed, but results were obtained for C1QTNF2. A few samples failed, because the DNA concentration was low.

Using the pyrosequencing assay, the methylation levels for the infected (Bps) and control mouse lung samples (PFS) were measured across the six CpG sites targeted in the promoter region of C1QTNF2. Higher methylation levels were found infected samples (Figure 2.3.9). Previous data showed that there is a significant decrease in methylation in the body region of C1QTNF2 (Figure 2.3.8 A) on infected human macrophages and C1QTNF2 expression was down-regulated in infected mice (Transcriptomics data obtained from the collaborators showed that C1QTNF2 had an expression fold change of -4.057 in Day 1 Bps-infected mice and -7.145 in Day 3 Bps-infected mice). These results align with suggestions in the literature that increased methylation in promoter regions is generally associated with down-regulation of expression, whereas methylation in gene body regions is positively correlated with gene expression [153,154].

(a)



(b)

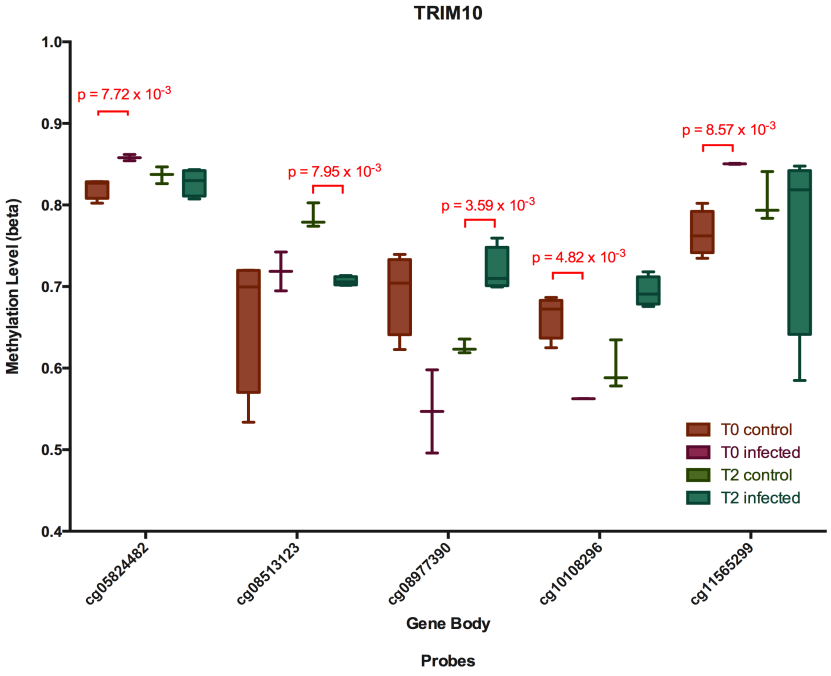


Figure 2.3.8. Methylation levels of genes C1QTNF2 (a) and TRIM10 (b) selected as targets for pyrosequencing methylation analysis on mouse lung samples. The x-axis lists the methylation sites associated with the gene and the y-axis is the normalized methylation level measured in Experiment 1 (U937 human macrophage).

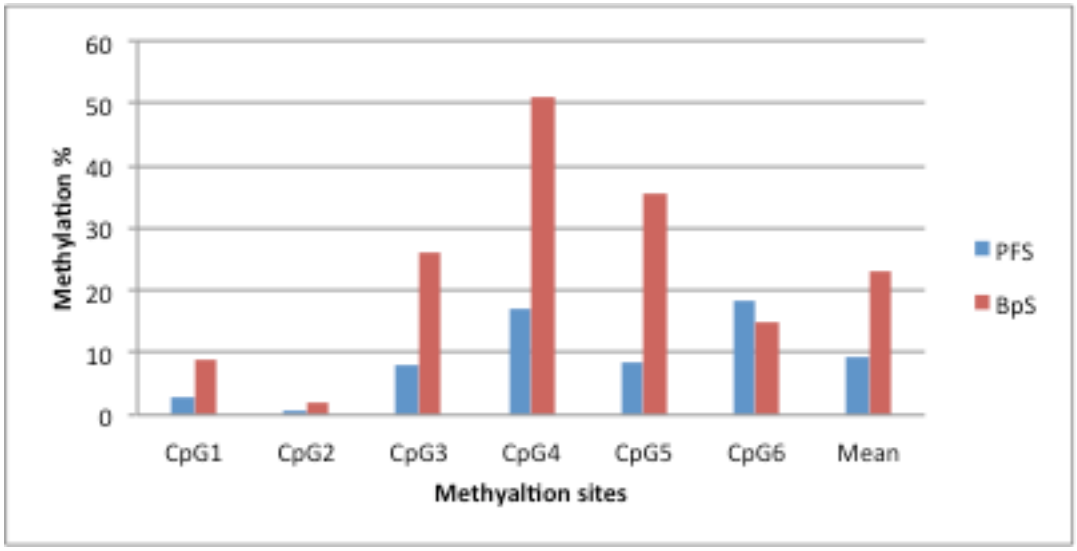


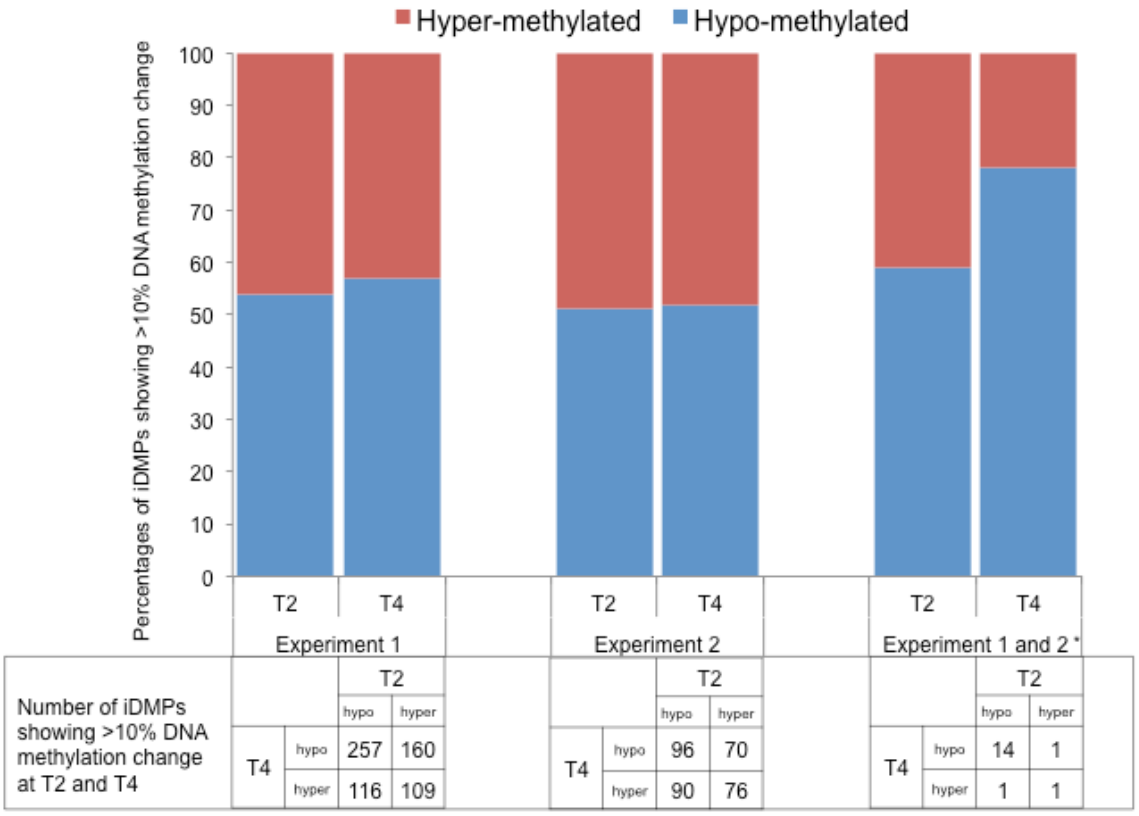
Figure 2.3.9. Pyrosequencing results on mouse lung samples. Target gene: C1QTNF2. BPS: *Burkholderia pseudomallei* infected; PFS: controls.

2.3.5 Experiment 2 - Differential methylation patterns

Next a replication experiment (Experiment 2) was performed following the same protocol as Experiment 1, but with additional sampling times and samples enabling to control for antibiotic effects. DNA and RNA were simultaneously collected at 1 hour (T1), 2 hours (T2), 3 hours (T3) and 4 hours (T4) post infection.

Reasoning that larger differences are potentially more biologically meaningful, infection-induced DMPs (iDMPs) were characterised by a greater than 10% DNA methylation difference between infected and uninfected cells. In Experiment 1, 10,279 iDMPs were identified (54% hypo-methylated; 46% hyper-methylated) at T2 and 4850 iDMPs (57% hypo-methylated; 43% hyper-methylated) at T4, with 642 iDMPs at both T2 and T4 (Figure 2.3.10). Using data from Experiment 2 a stringently filtered dataset of 388 iDMPs was obtained

as conserved between the two experiments, which were used for subsequent analysis. These DNA methylation changes were highly correlated across both experiments (T2: r=0.83, p<0.0001 – see Figure 2.3.11a; T4: r=0.73, p<0.0001 – see Figure 2.3.11b). 264 iDMPs were conserved at T2, 141 iDMPs were conserved at T4 and 17 iDMPs were conserved at both T2 and T4.



Number of iDMPs showing >10% DNA methylation change at T2 and T4

		T2 hypo	T2 hyper
T4	hypo	257	160
T4	hyper	116	109

		T2 hypo	T2 hyper
T4	hypo	96	70
T4	hyper	90	76

		T2 hypo	T2 hyper
T4	hypo	14	1
T4	hyper	1	1

Figure 2.3.10. Quantification of infection-induced DMPs (iDMPs). * iDMPs conserved between the two experiments.

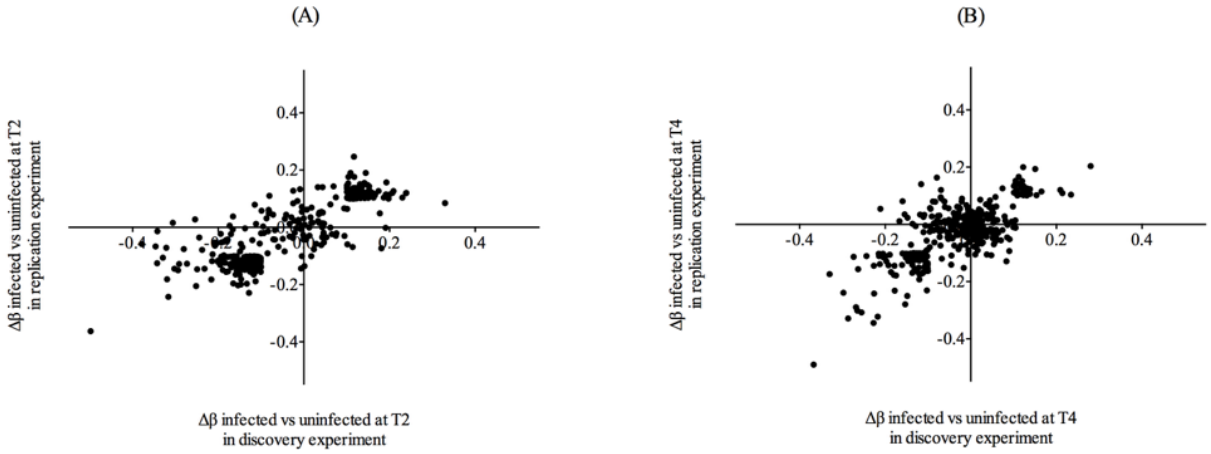


Figure 2.3.11. Methylation changes in experiments 1 and 3 at replicated iDMPs. (a) Changes at T2. (b) Changes at T4.

The methylation levels of the uninfected samples were plotted in a heatmap (Figure 2.3.12). This shows that differential methylation occurs in heavily, partially and sparsely methylated regions. When the methylation levels of infected samples are compared to uninfected samples (Figure 2.3.13), in general, there are hypo-methylated sites.

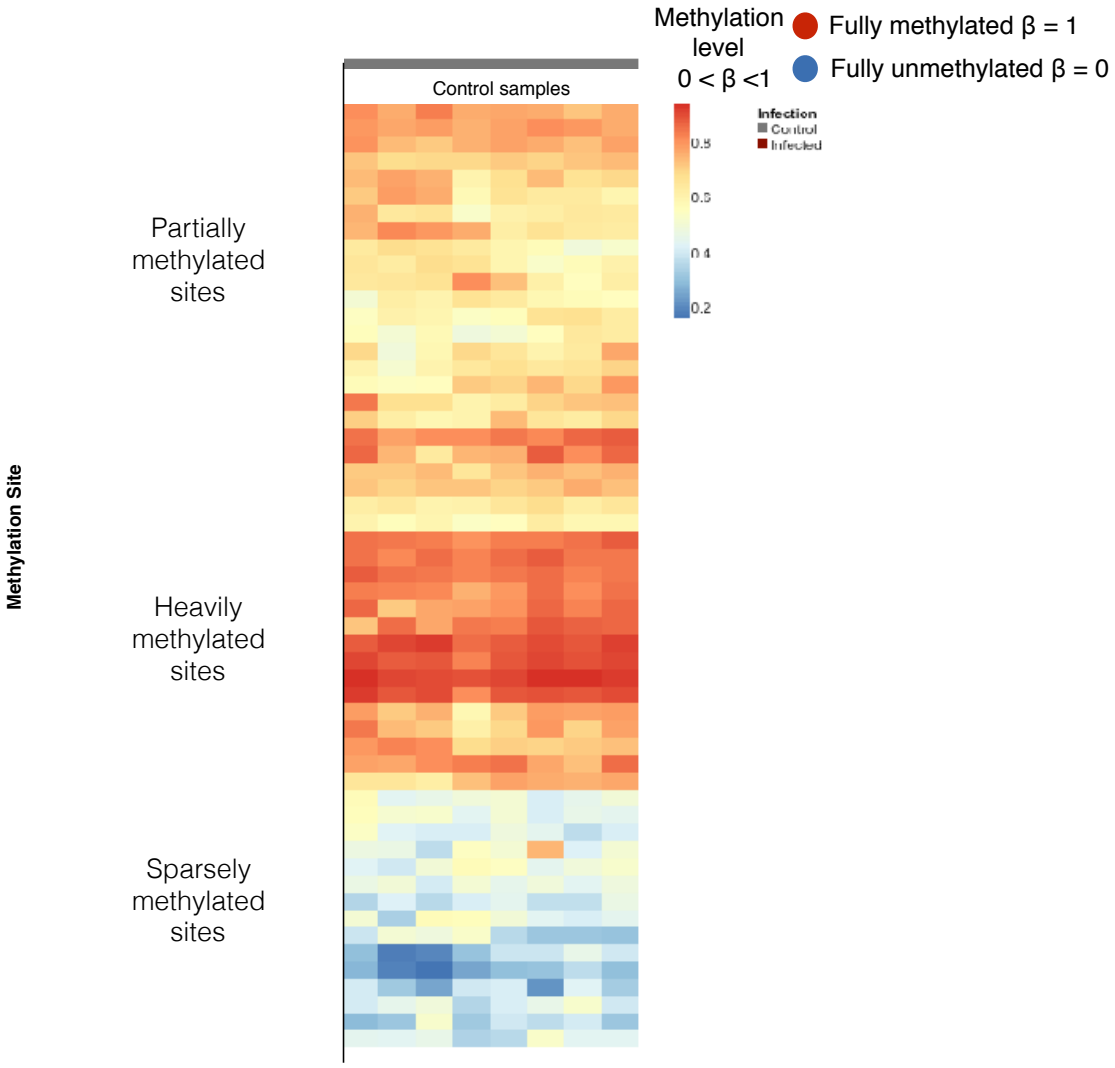


Figure 2.3.12. Heatmap of methylation levels of control samples at iDMPs.

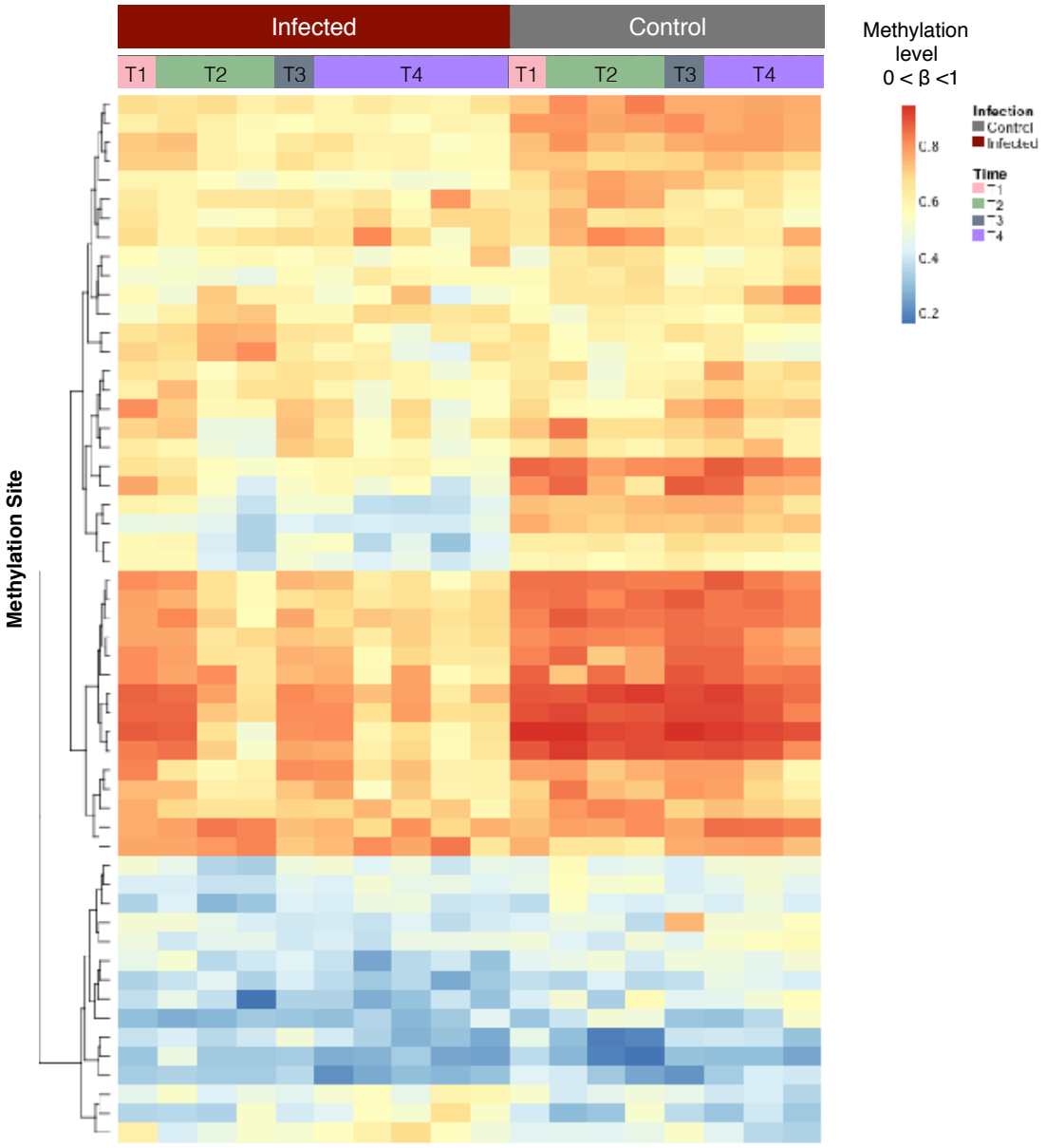


Figure 2.3.13. Heatmap of methylation levels of infected and control samples at iDMPs.

In addition to quantifying DNA methylation at T2 and T4 in experiment 1 and 2, we also profiled samples collected at T1 and T3 in experiment 2. In experiment 2, only one sample was collected for each group. As in both experiments measurements have been taken at T2 and T4, DMPs identified at

T2 and T4 in the first experiment were compared to the corresponding measurements at the same time point in the second experiment. Validated probes were assigned to categories based on the change in score at successive time points.

Several distinct patterns of DNA methylation across the four time points were observed at the conserved iDMPs (Table 2.3.3.). As there is more than one sample for each group in experiment 1, the standard deviations of the methylation levels were represented by error bars. Signatures with more than 5% differential methylation at all time points were classified as, constant hyper-(+5%) or constant hypo-methylated (-5%) probes. Signatures with more than 5% differential methylation at early time points (T1, T2, T3) and no change at the T4 time point were classified as early response probes. Signatures with no response at the T1 time point and with more than 5% differential methylation at later time points were classified as late response probes. Signatures with no response at T1 and T4 and more than 5% differential methylation at T2 or T3 were classified as transient responses. The remainder of the patterns was classified as oscillatory probes.

21 probes (5.41%) showed consistent changes in DNA methylation across all time points in both experiments (for example cg17676428 (Fig 12a) and cg21823425 (Figure 2.3.14. b)). 23 probes (5.93%) showed large changes at the early time-points that diminished at the later time-points (for example cg15470658 (Figure 2.3.14. c) and cg11630696 (Figure 2.3.14. d)). 55 probes (14.18%) displayed a lag in response, with DNA methylation changes only occurring at the later time points after infection (for example cg14113958 (Figure 2.3.14. e) and cg06284169 (Figure 2.3.14. f)). 99 probes (25.52%) showed a transient response (for example cg14173033 (Figure 2.3.14. g) and

cg00782124 (Figure 2.3.14. h)). Finally, 190 probes (48.97%) had an oscillatory response (for example cg00400832 (Figure 2.3.14. i) and cg15392054 (Figure 2.3.14. j)). The losses of methylation were more prominent compared to gains in DNA methylation patterns, especially in iDMPs exhibiting a constant response. Immune processes should be tightly regulated to avoid excessive damage to the host. Inflammatory processes, for example, are necessary to initiate a protective response against harmful bacteria. However, prolonged and excessive inflammation can damage bystander normal tissue [155]. The patterns demonstrated in Figure 2.3.14 allow evaluation of the role of DNA methylation in the kinetics of immunoregulation. Components associated with early DNA methylation response can play role in eradicating bacteria. Late responses can terminate processes such as inflammation after the bacteria have been cleared. Relative to the burden of infection certain responses are elevated or dampened at different times. Transient and oscillatory responses could point the time to key events during infection with *B. pseudomallei*. Polymerisation of the host actin, for instance, is a process that provide bacteria motility in the host cytosol after bacteria evade host innate response [156]. Methylation patterns mapping to the time course of host responses can decipher characteristic features of *B. pseudomallei* in their ability to infect host cells.

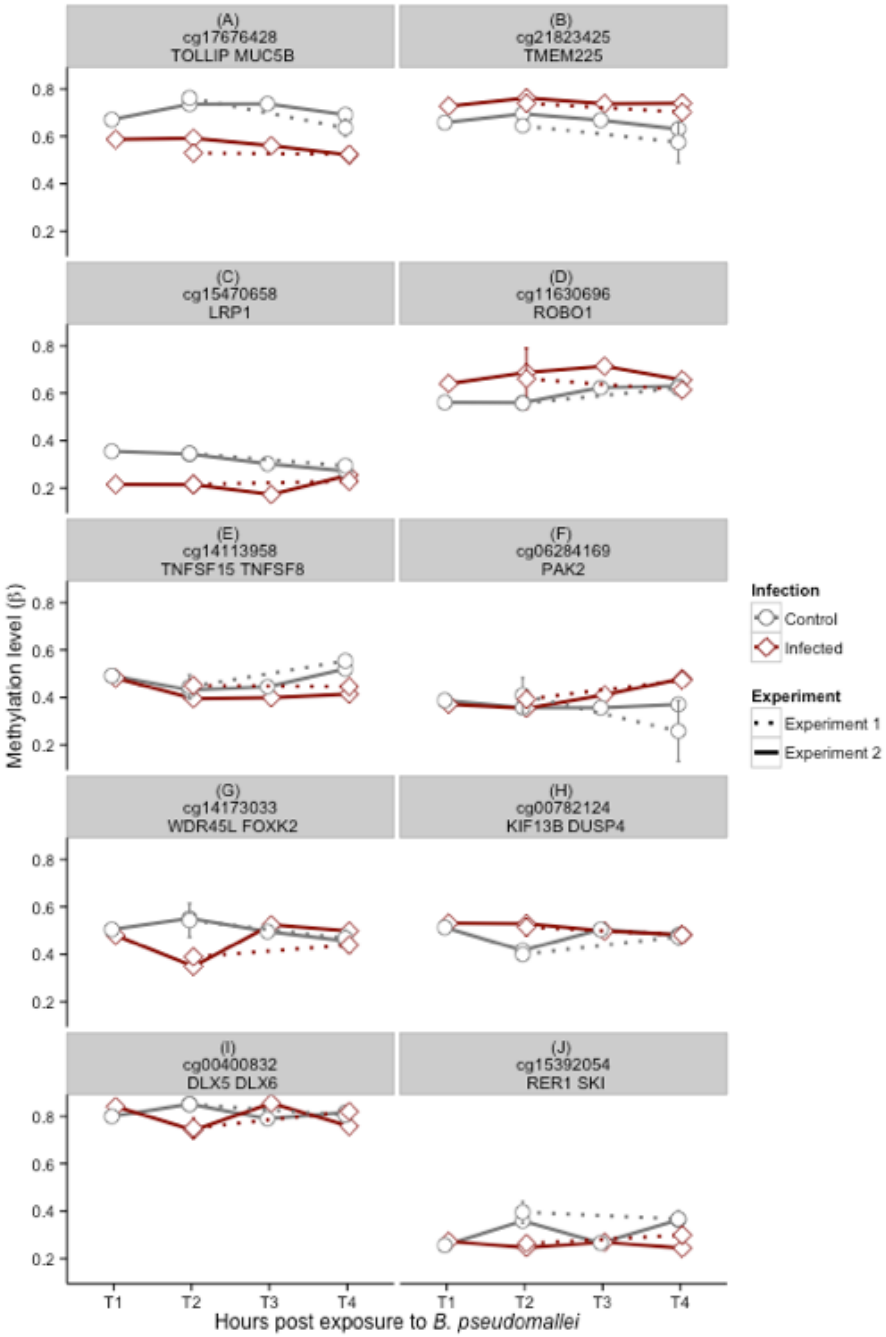


Figure 2.3.14. Differential methylation patterns. Each panel shows a representative gene, with a particular temporal methylation pattern, as discussed in more detail in the text. The error bars represent the standard deviation of the methylation levels of the first experiment samples: (A) constant hypo-methylation; (B) constant hyper-methylation; (C) early response hypo-methylation; (D) early response hyper-methylation; (E) late response hypo-methylation; (F) late response hyper-methylation; (G) transient response hypo-methylation; (H) transient response hyper-methylation; (I) & (J) oscillatory response.

Table 2.3.3. Number of iDMPs assigned to categories based on differential methylation patterns. Signatures with differential methylation at all time points (1 hour (T1), 2 hours (T2), 3 hours (T3) and 4 hours (T4) post infection) were classified as constant hyper- or constant hypo-methylated probes. Signatures with differential methylation at early time points (T1, T2, T3) and no change at the T4 time point were classified as early response probes. Signatures with no response at the T1 time point and with differential methylation at later time points were classified as late response probes. Signatures with no response at T1 and T4 and differential methylation at T2 or T3 were classified as transient responses. The remainder of the patterns were classified as oscillatory probes.

Category	number of probes	hyper-methylated	hypo-methylated
A: constant response	21	3	18
B: early response	23	10	13
C: late response	55	11	44
D: transient response	99	42	57
E: oscillatory response	190	NA	NA

In experiment 2, additional samples were collected. These were DNA from uninfected cells not treated with the antibiotic kanamycin. For the set of iDMPs no differences were found between the uninfected cells treated with kanamycin and the uninfected cells without kanamycin.

To explore the biological significance of the genes mapping to the set of iDMPs, genes were annotated with gene ontology (GO) terms and searched for over-representation of categories using the package GOseq, which weights genes based on the number of probes per gene. Enriched biological functions were identified using semantic similarity, which provides a measure of functional similarity (Figure 2.3.15). Functionally similar GO terms appear closer in the plot. Functions related to cellular responses to stress can point to the host cell detecting the presence of bacteria. Interleukin related components play a role in

stimulating immune responses. This can show that DNA methylation processes can initiate or be a result of inflammation. The regulation of adhesion molecules can suggest that the bacteria is gaining motility in the host cell through manipulating the host DNA.

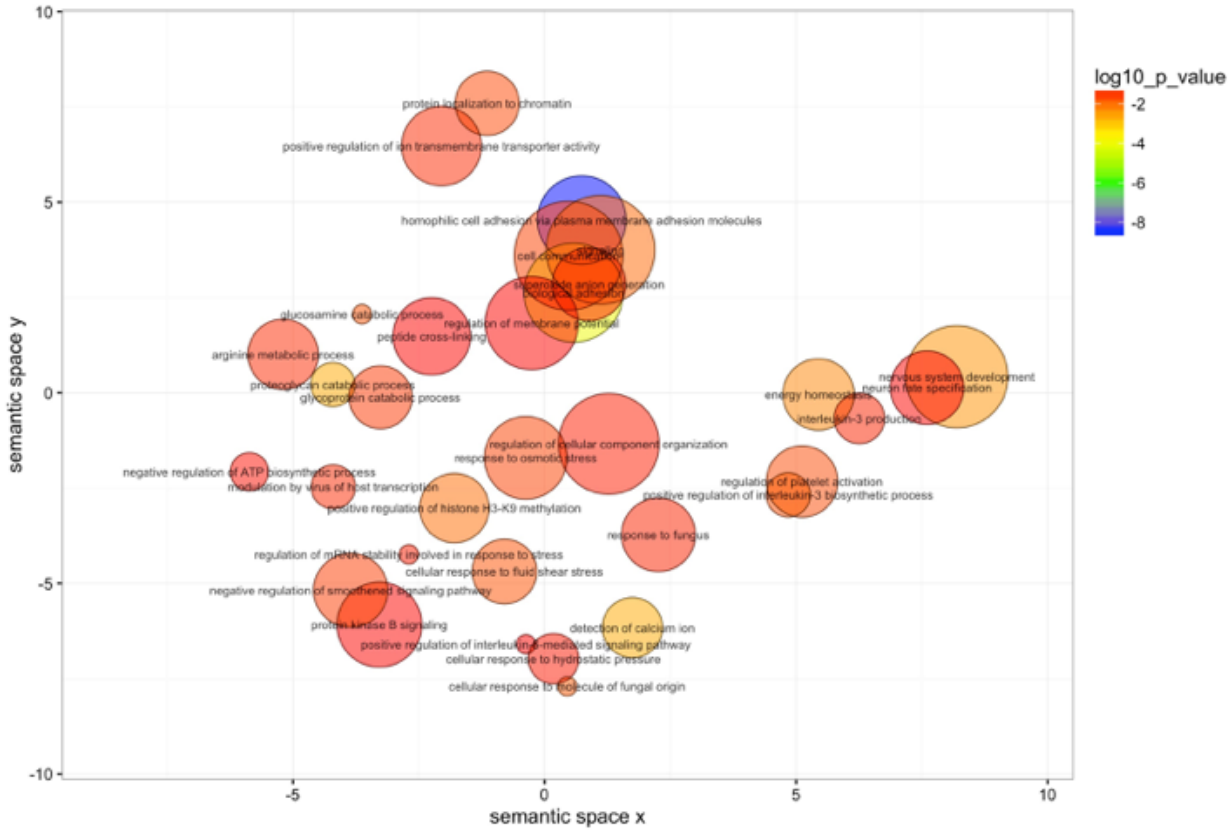


Figure 2.3.15. Gene ontology terms enriched (p<0.05) in genes mapping to conserved iDMPs. The colour scale represents the *p*-values calculated using GOseq. The non-redundant gene ontology terms are clustered using REVIGO.

2.3.6 An alternative approach for identifying differential methylation patterns: Multi-scale Gaussian Mixtures for identifying differentially methylated sites

Given the small sample size of the study, it is crucial to develop the right approach to detect the differentially methylated probes. So far, the measure to assess methylation changes was the effect size, the magnitude of methylation

level differences (Δβ) between the infected and the control samples. Assessing the significance of these changes is also important in identifying DMPs. In this section, the multi-scale Gaussian mixtures method was used to identify DMPs, based on statistical significance. This method was proposed by Yang *et al.* [157] to extract homogeneous and heterogeneous gene expression patterns across species. The sum and difference of expressions are modelled using multi-scale Gaussians. In this section, the multi-scale Gaussian model was applied to reveal the differential methylation patterns between the infected and control cells.

The model is composed of two Gaussians, shown as red and blue curves in Figure 2.3.16, which have identical means, but different variances. The Gaussian with the smaller variance is referred as the null density (red) and the Gaussian with the larger variance (blue) as the alternative density. A Bayesian learning process is initiated for each difference vector and the null and alternative densities are estimated after convergence is achieved. For each methylation site, the probabilities that it belongs to the null or alternative distribution are calculated. In the histogram of the methylation sites in Figure 2.3.16, sites with alternative probability greater than the null probability are defined as the differentially methylated sites (these correspond to the area outside the red curve under the blue curve).

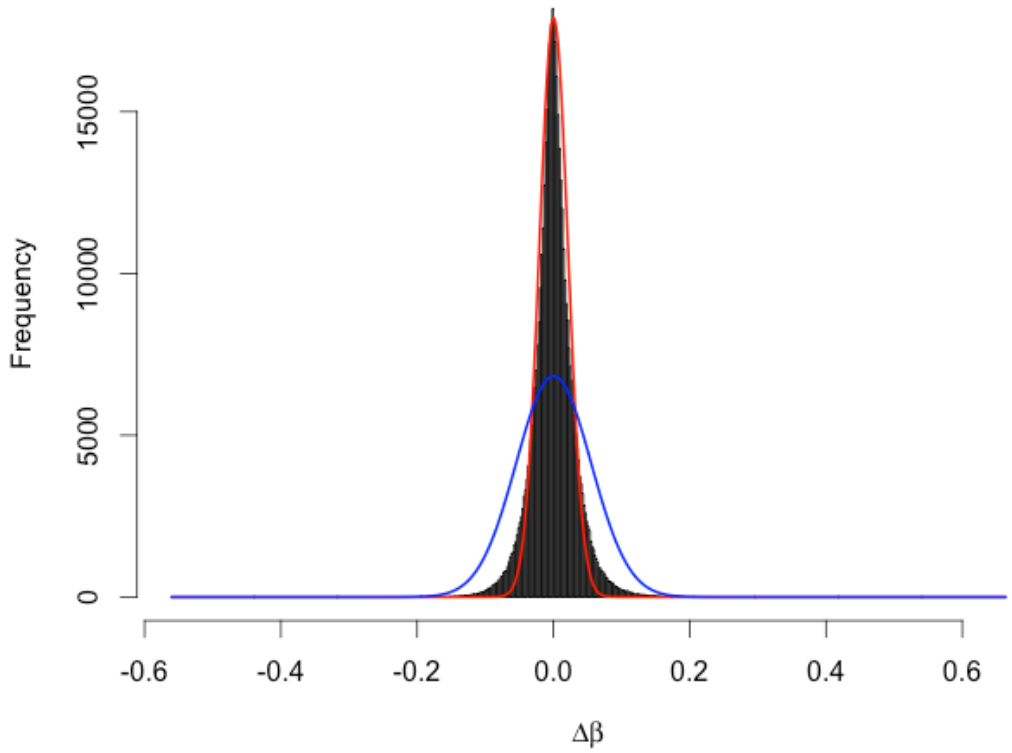


Figure 2.3.16. Histogram of methylation level differences modelled with a Multi-scale Gaussian Mixture Model.

Using the Multi-scale Gaussian Mixtures method on the methylation data presented in this chapter, 5269 methylation sites were obtained, which are the top 1% of the sites ranked by their posterior probability. Figure 2.3.18 is a heatmap of all the differentially methylated sites presented as a Venn diagram (Figure 2.3.17) when the sites are seriated by their methylation difference. This reveals patterns such as demethylation at T1, methylation at T2 and T3, demethylation at T4 etc. The results obtained with the Multi-scale Gaussian Mixtures method were similar to the results from previous analyses where the focus was on the probes that have big methylation differences. Even though same DMPs were identified, using the Multi-scale Gaussian Mixtures method a significance measure was assigned to these DMPs. This verified that the big differences identified previously are statistically significant.

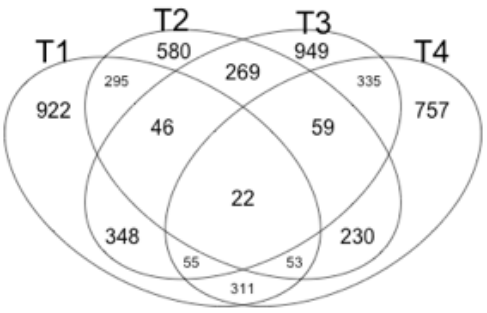


Figure 2.3.17. Venn diagram showing the number of differentially methylated sites across the time points (top 1% ranked by posterior probability).

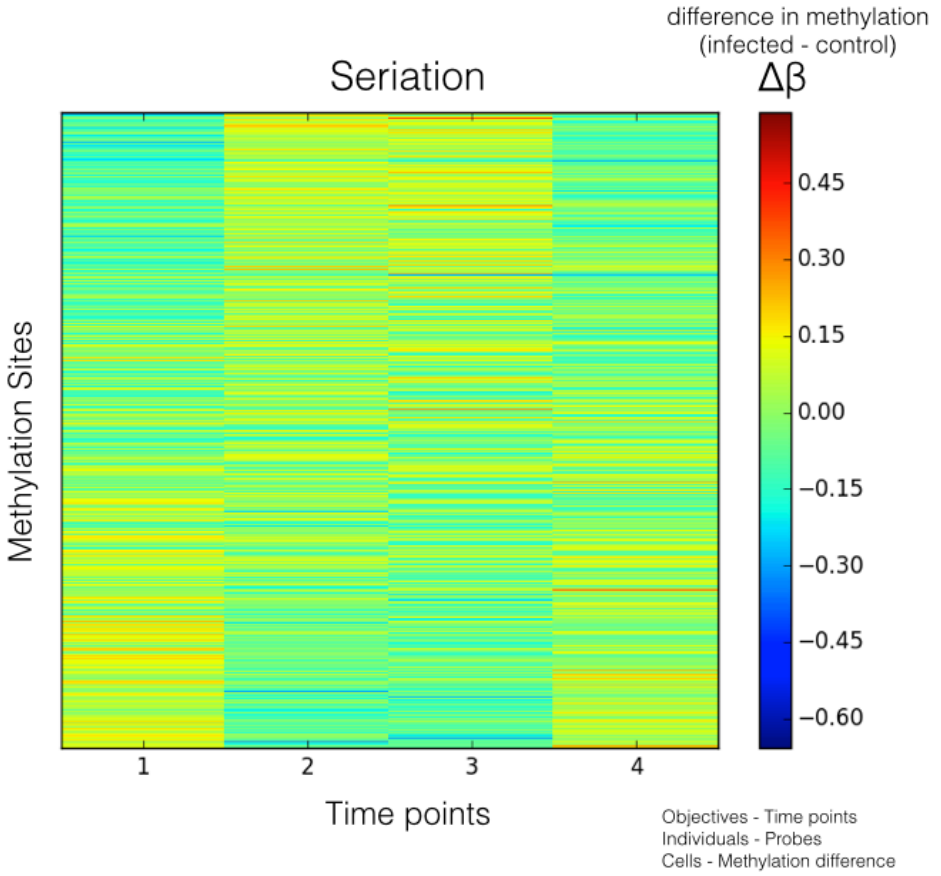


Figure 2.3.18. Heatmap showing the Δβ values for differentially methylated sites across the time points.

2.3.7 Effect of differential methylation on gene expression

2.3.7.1 Measuring expression levels using qPCR

In order to establish whether the altered methylation patterns produced corresponding changes in expression levels, the expression levels of two genes (*TOLLIP*, *DDAH2*) were quantified using quantitative real time PCR. These genes were selected from the group of constant response iDMPs. mRNA levels in infected and uninfected samples. At T1 and T2, neither of the selected genes showed a significant change in expression levels (Fig 2.3.19). *DDAH2* showed elevated expression levels in infected cells at T3 and T4, *TOLLIP* showed elevated expression levels in infected cells at T4. Both genes thus exhibited a delayed effect on gene expression following the alteration of DNA methylation. This effect was inversely correlated with the Δβ values at the differentially methylated sites.

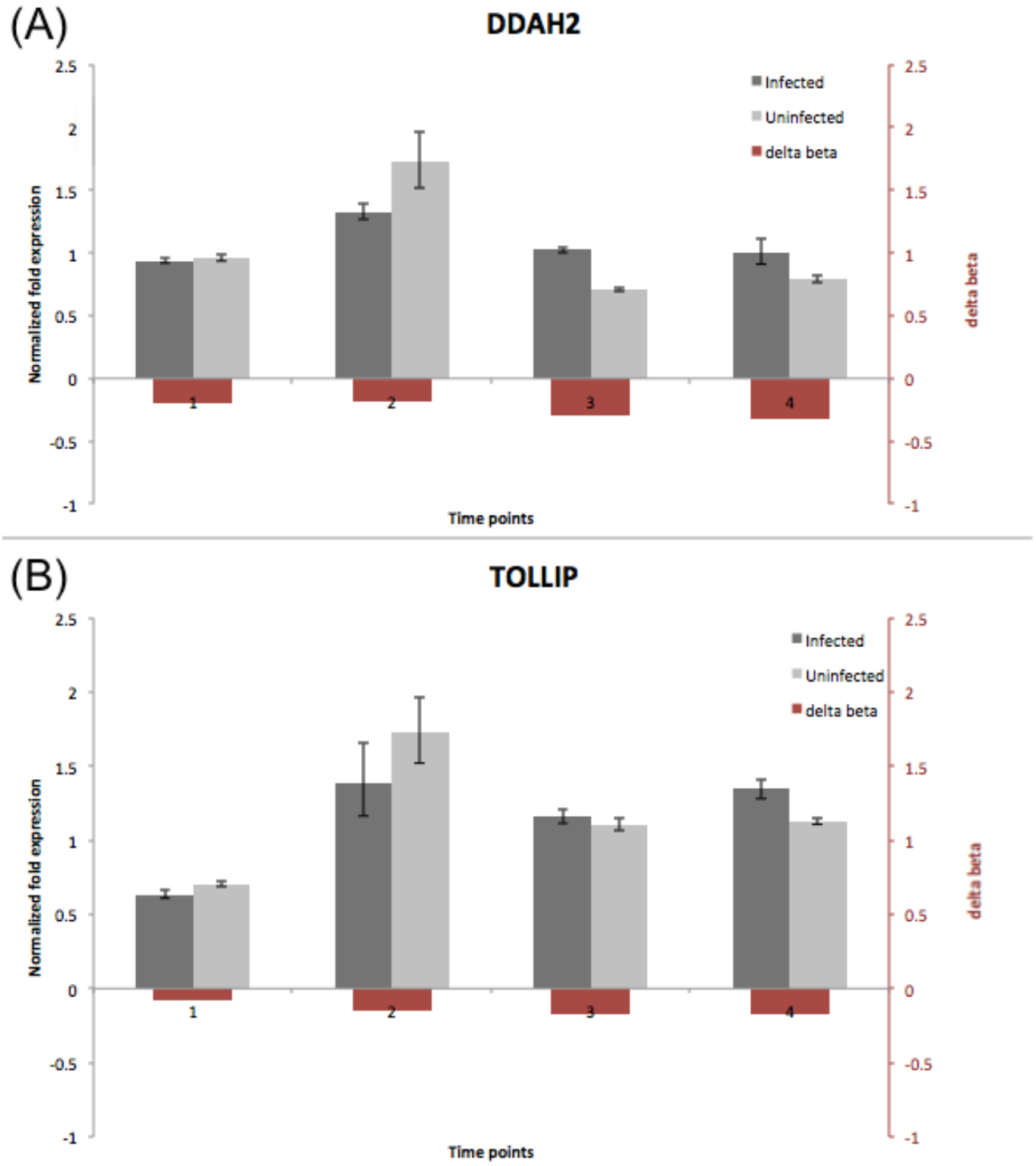


Fig 2.3.19. Gene expression levels following altered methylation patterns for DDAH2 (A) and TOLLIP (B) in infected and uninfected cells. Red bars: Δβ values. Grey bars: fold expression.

2.3.7.2 Comparison with publically available transcription data

The iDMPs were compared with transcriptomic data from a previous study of *B. pseudomallei* infection in humans which identified 2604 human genes, that were differentially expressed in patients with septicemic melioidosis [158]. Of these, 76 genes were annotated to the iDMPs we identified in our study. These included genes involved in immune system processes (*BCL11B*, *CDKN1C*, *GLMN*, *HLX*, *IL1R2*, *IRF8*, *MAEA*, *MEF2C*, *MR1*, *NBEAL2*, *PRKCH*, *PTGDR2*, *STK3*, *TNFSF8*, *TRIM27*), response to stress (*ADRB2*, *APBB1IP*, *DTNBP1*, *FBXO31*, *MARCH1*, *MSRA*, *PKD2*, *SCARB1*, *ZMYND11*), and inflammatory response (*CD44*, *HDAC4*, *HIF1A*, *IL18RAP*, *TOLLIP*, *IER3*, *NT5E*).

2.3.8 Comparison with publically available DNA methylation changes during infection with other pathogens

Innate immune cells, such as macrophages or dendritic cells, are recruited in response to pathogens to initiate defense mechanisms. While this chapter explored DNA methylation changes in macrophages, a recent study identified epigenetic regulation in human dendritic cells before and after *in vitro* infection (18 hours post infection) with *Mycobacterium tuberculosis* [80]. The proportion of iDMPs compared to the total number of sites probed was comparable to that in this chapter (0.000912 in this chapter; 0.000649 in the *M. tuberculosis* study). There weren't any stringently filtered iDMPs which were identical in cells infected with *B. pseudomallei* or with *M. tuberculosis*. However, when we considered the genes to which these iDMPs mapped, we identified 121 genes (median distance of ~95 kb from the nearest transcription start site) that showed differential methylation in both studies. These included genes

involved in T cell responses (*BCL11B*, *FOXO1*, *KIF13B*, *PAWR*, *SOX4*, *SYK*), actin cytoskeleton organization (*ACTR3*, *CDC42BPA*, *DTNBP1*, *FERMT2*, *PRKCZ*, *RAC1*), and cytokine production (*FOXP1*, *IRF8*, *MR1*) (Fig 2.3.20).

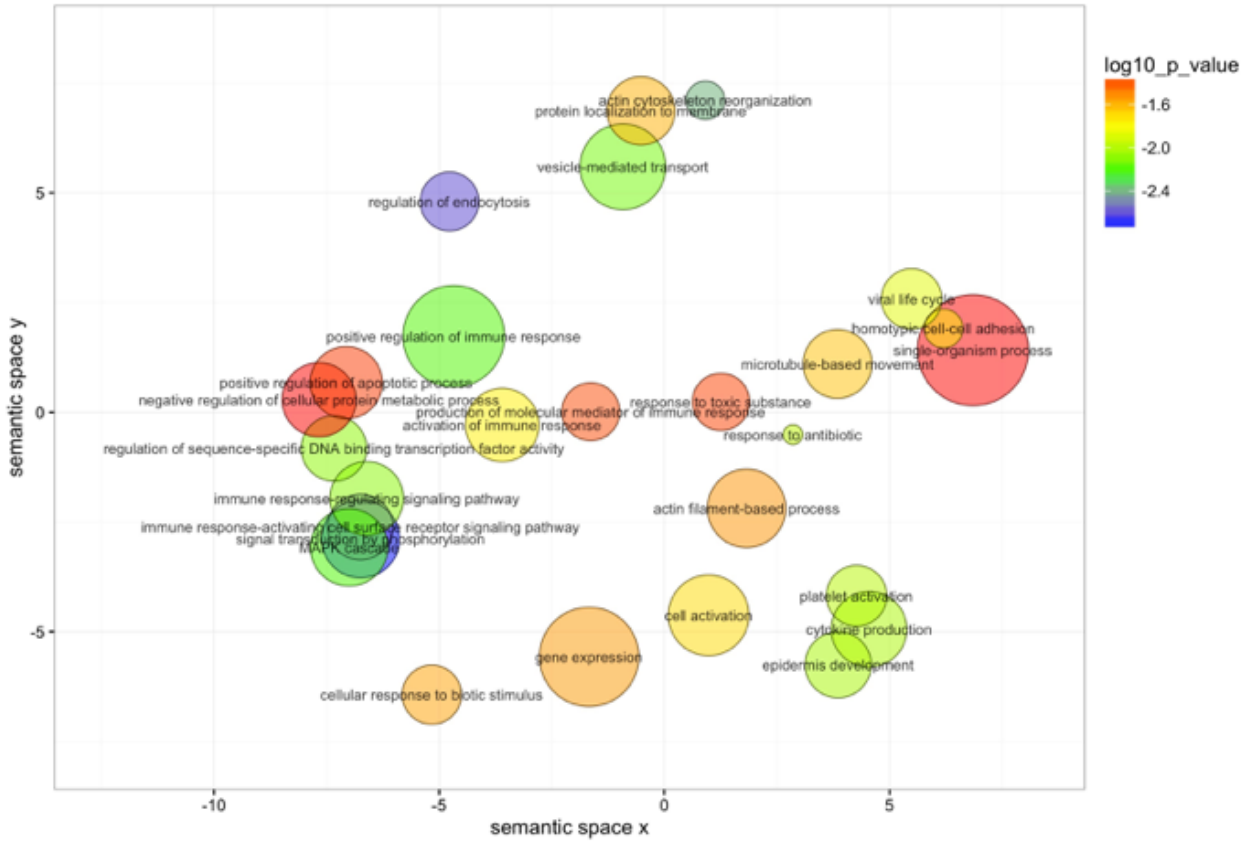


Figure 2.3.20. Gene ontology terms enriched (p<0.05) in genes annotated to the iDMPs in *B. pseudomallei* and M. tuberculosis infections. The colour scale represents the *p*-values calculated using GOseq. The non-redundant gene ontology terms are clustered using REVIGO.

Another recent study examined DNA methylation changes in human macrophages infected with *L. donovani*, an intracellular parasite capable of regulating the host defense (Marr *et al*). This study reported 733 probes with significant methylation changes between infected and uninfected cells [19]. Of these, 4 were differentially methylated in our study, at all time points (Table 2.3.4). These probes, which were all hypo-methylated, were in the vicinity of the *TOLLIP*, *DIP2C*, *DAGLB* and *SFRS16* genes. These genes play important roles

in regulating the innate immune response (*TOLLIP* and *DAGLB*) or transcription (*DIP2C* and *SFRS16*).

Table 2.3.4. CpG sites with differential methylation in live infected versus uninfected human macrophages in *L. donovani* infection [19] and in *B. pseudomallei* infection (this study).

		***L. donovani* infection**	***B. pseudomallei* infection**			
CpG site	**UCSC RefGene name**	**Δβ**	**Δβ at T1**	**Δβ at T2**	**Δβ at T3**	**Δβ at T4**
cg17676428	TOLLIP	-0.1326	-0.0839	-0.2025	-0.1765	-0.1304
cg26561082	DIP2C	-0.0837	-0.0407	-0.1724	-0.2340	-0.1897
cg22741272	DAGLB	-0.0871	-0.0400	-0.1170	-0.1343	-0.1178
cg26985354	SFRS16	-0.1000	-0.0609	-0.1734	-0.2026	-0.1717

2.4 Discussion

This chapter demonstrates that there are reproducible significant changes in host macrophage DNA methylation, at multiple CpG sites in the host cell genome, following infection with *B. pseudomallei*. The results not only provide support for the importance of DNA methylation in understanding infection, but also reveal the dynamics of these changes during the early stages of infection.

Infection induced differentially methylated probes (iDMPs) occur in heavily, partially, and sparsely methylated regions. Most of the methylation changes in infected cells are losses rather than gains in methylation. This trend is in good agreement with previous studies on cells infected with other pathogens. Indeed, similar hypo-methylation motifs were seen during *M.*

tuberculosis infection [80]. Furthermore Marr *et al*. reported a larger proportion of hypo-methylated CpG sites in *L. donovani* infected macrophages [81]. Comparison of *M. tuberculosis* and *B. pseudomallei* induced changes in methylation patterns revealed that although a number of common genes were differentially methylated, there were no conserved iDMPs. However, neither presence nor lack of correlation can be interpreted too strongly, due to the different complexities associated with each dataset. These studies used different cell types and hence we cannot discount the possibility that there would be conserved iDMPs if the experimental conditions were identical.

iDMPs showing the greatest changes in DNA methylation were found to be in the vicinity of genes involved in inflammatory responses, intracellular signalling, apoptosis and pathogen-induced signalling. Toll-interacting protein (*TOLLIP*) plays an important role in the regulation of innate immunity and therefore with early responses to pathogens. Increased expression of *TOLLIP* has been reported in humans infected with *B. pseudomallei* [5]. The precise role of *TOLLIP* in protective immunity is still being clarified. One function of the protein is to interact with *IL1RI*, *TLR2* and *TLR4* after lipopolysaccharide (LPS) activation, thereby modulating the NF-KB and JNK signaling pathways [20]. The role of *B. pseudomallei* LPS in virulence [21] and the activation of *TLR2* and *TLR4* signalling during *B. pseudomallei* infection is well documented [22-24]. It is possible that the differential methylation of *TOLLIP* we have seen modulates the early immune response to *B. pseudomallei* LPS. The finding that *TOLLIP* was also differentially methylated in *L. donovani* infected cells may reflect the other roles of this protein in regulating other parts of the innate immune response [20]. Further work is required to investigate whether differential

methylation of *TOLLIP* occurs following infection with a broader range of prokaryotic and eukaryotic pathogens.

A previous study found that the alternative pathway of macrophage activation is activated after infection with *Leishmania*, and that this is characterised by the induction of genes including *DAGLB* (encoding diacylglycerol lipase beta) [25]. The roles of the classical and alternative pathways of macrophage activation in controlling *B. pseudomallei* infection are not known. However, the pattern of growth of *B. pseudomallei* in cells we have seen is consistent with macrophage activation. The findings presented in this chapter lend weight to the suggestion that this protein plays a role in the pathogen-common pathways of macrophage activation [25].

Two of the enriched functions identified in this chapter are cell adhesion and cell communication, categories that include genes involved in actin related functions and programmed cell death. *B. pseudomallei* is known to promote the polymerisation of host actin at one pole of the bacterium in an *ARP2/3*-dependent manner [26]. This enables motility of the bacteria and is essential for the uptake of bacteria into cells and the spread of bacteria within and between cells. Gene ontology results presented in this chapter demonstrated that the differential methylation of several genes are involved in actin binding or polymerisation. These genes included *CDC42BPA* a component of the BLOC-1 complex, which can interact with WASH - a regulator of *ARP2/3* [27] and *ACTR3*, a major component of the *ARP2/3* complex [28]. Differential methylation was also found for one of the central regulators of actin polymerisation, the *PAK1* p21-activated kinase [29].

A range of genes associated with the immune response to infection were differentially methylated. These include genes encoding cytokines, chemokines

and their receptors (*CCR4, IL1R2, IL18RAP, TNFSF15, TNFSF8, CCL28*) and signalling pathways (*CASP8AP2, TOLLIP, SYK, ZBP1, MAP4K4, MBIP*). *B. pseudomallei* is an intracellular pathogen, and professional phagocytes are believed to be an important niche for growth and survival in the host [30,31]. Host interferon gamma plays a key role in the control of *B. pseudomallei* infection [32,33] and suppression of the host innate response to *B. pseudomallei* infection has previously been attributed to downregulation of the type I interferon gamma signalling pathway by the bacterial effector TssM [34]. Previously, $CD44^{high}$ CD8 T-cells have been shown to be an important source of interferon gamma [35] and this chapter presents evidence of differential methylation of CD44. There were also evidence of differential methylation of a number of genes (*IRF-8, IFNE and ZBP1*), which are involved in the regulation of interferon gamma expression. These findings provide new insights into the possible molecular mechanisms underpinning this suppression.

The *DDAH2* gene encodes an enzyme in the nitric oxide generation pathway and it is known that nitric oxide has potent antibacterial activity against *B. pseudomallei* [36,37]. The differential methylation of the gene encoding the *NOX4 NADPH* oxidase that we have seen, might explain the activation of this enzyme in phagocytes infected by the bacterium [38].

Another important feature of disease caused by *B. pseudomallei* is the ability of the pathogen to establish chronic or persistent infections. This suggests that *B. pseudomallei* is able to subvert the host immune response to infection. Our results provide new insights into how the host cell functions are modulated. The differential methylation of genes involved in memory T-cell responses, such as *CD44*, *FOXO1* and *FOXC1*, might contribute to the inability of the host to mount responses capable of clearing infection.

We also found differential methylation of genes associated with ubiquitination of proteins targeted for degradation, including the *SPSB4* and *WSB1* gene products which are associated with substrate recognition, and the *NEDD4L* and *SIAH1* genes encoding ubiquitin ligases. The deamidation of the *NEDD8L* protein by a *B. pseudomallei* type III effector protein (*CHBP*) has previously been demonstrated, and this triggers apoptosis of host cells [39]. However, the modulation of host cell ubiquitination has also been associated with the suppression of host immunity [40], and the downregulation of the NF-kappaB/type I IFN signalling pathways has been attributed to the ubiquitination of signalling molecules [34]. The findings in this chapter demonstrate that methylation may contribute to modulation of the ubiquitination pathway.

In summary, this chapter provides new insight into the extent to which infection with a bacterial pathogen results in differential methylation of host cell DNA. The findings indicate that differential methylation of host DNA occurs on a much greater scale than previously suggested. This would certainly raise interest in investigating DNA methylation during infection with other pathogens. To unravel the most accurate signatures of infection, it is important to determine whether these methylation changes are a general host response to infection or pathogen-specific responses. Whether differential methylation is a consequence of the direct interaction of bacterial methyltransferases with host cell DNA, or the consequence of modification of host cell methyltransferase activity awaits investigation. The findings presented in this chapter can serve to develop DNA methylation based biomarkers for diagnosing melioidosis. These biomarkers are more specific and sensitive than commonly used protein biomarkers [159]. Therefore they have great potential to be used in clinical diagnostic tests and in the development of specialised drugs.

CHAPTER 3. HOST DNA METHYLATION CHANGES DURING MELIOIDOSIS: IN VIVO INFECTION MODEL

3.1 Introduction

Because of its nonspecific and diverse clinical presentation, melioidosis is often hard to diagnose. A variety of antibiotics are ineffective in treating melioidosis and the disease has a high mortality rate. Studies emphasise the importance of early diagnosis in appropriate treatment and improved disease outcome [160].

There is growing recognition for the benefits of identifying biomarkers specific to infection in accelerating the development of diagnostic assays and in obtaining faster and more accurate diagnostic tests. Several studies have aimed to identify blood-based biomarkers for septicaemia [161–163], which is the most common presentation of melioidosis. Sepsis represents a pattern of response by the immune system to infection. Pro- and anti-inflammatory cytokines and chemokines have been found to play a role in sepsis. Recent studies employed whole blood transcriptome profiling to unravel the regulation of the immune system during sepsis [161–163]. These studies revealed potential transcriptional responses in sepsis, including signatures distinguishing septicemic melioidosis from sepsis caused by other infections [124].

Despite these efforts, the pathogenesis of sepsis is not yet fully understood. Recently, the investigation of the epigenomic information in blood, particularly DNA methylation, emerged as a means of understanding the molecular mechanisms underlying human health and disease. DNA methylation signatures in blood have been identified for a wide range of other diseases [164–167].

The motivation behind this chapter is that identifying DNA methylation differences associated with complex infectious diseases will improve clinical diagnosis. To this end, this chapter investigated the association of whole blood DNA methylation with melioidosis. Analysis of DNA methylation in melioidosis patients can facilitate development of accurate biomarkers for detection of *B. pseudomallei* infection. These DNA methylation-based biomarkers can signal disease-specific sites, which cannot be detected using transcriptome profiling.

3.2 Methods

3.2.1 Samples demographic

For this study, whole blood samples of 41 individuals, who were part of a larger study at Khon Kaen University, were recruited. Signed, written consent was obtained from all of the participants and the project was approved by the ethics committee. DNA samples were prepared under the supervision of Prof. Ganjana Lertmemongkolchai at Khon Kaen University. The samples consisted of 28 control subjects without a history of melioidosis and 13 patients diagnosed with melioidosis sepsis. Sepsis is defined using the same criteria outlined by Pankla *et al.* [168], namely presentation with two or more of the following criteria for the systematic inflammatory response syndrome: fever (temperature >38°C or <36°C), tachycardia (heart rate >90 beats/minute), leukocytosis or leukocytopenia (white blood cell count ≥12 x 10^9/l or ≤4 x 10^9/l). Anonymised DNA samples were shipped to the UK with additional patient data. This data included the following: age, gender, race, source of infection and underlying disease(s).

3.2.2 DNA methylation analysis

Genome wide methylation levels of infected and uninfected samples were profiled using using the Infinium HumanMethylation450 BeadChip (450K) (Illumina Inc.) as explained in Chapter 2. Briefly, DNA was sodium bisulfite converted using the EZ-96 DNA Methylation kit (Zymo research, CA, USA). Post-hybridisation allelic-specific single-base extension of the probes incorporates a fluorescent label enabling detection. Red and green channel intensity files were obtained for each sample in the. idat file format.

3.2.3 Data acquisition and normalisation

All quality control checks were performed using R (version 3.2.3). Signal intensities were imported into R using the *methylumi* package. For each sample, the distribution of methylated and unmethylated signals were plotted to check whether there were any extreme outliers.

For each CpG site a β-value was calculated as explained in Section 2.2.3. Quality control checks and quantile normalisation were implemented using the *dasen* function of the *wateRmelon* package [136]. Samples with more than 1% of sites with a detection *p*-value greater than 0.05 were removed, as were probes with 1% of samples with a detection *p*-value greater than 0.05. Probes were removed if they had a bead count less than 3 in 1% of samples. Further problematic or uninformative probes were filtered out. These were non-specific CG probes (those without a proper Illumina identifier, cgXXXXXXX), probes with common SNPs (Minor allele frequency (MAF) >0.05 in an Asian population) in the single (or multiple) bp extension of the CpG site (a ±10bp cut-off was used) [169], and cross-hybridising probes [138,170]. Finally, probes on sex chromosomes were removed, leaving 394048 probes for further analysis.

3.2.4 Unsupervised analysis

To detect patterns in methylation data, hierarchical clustering and principal component analyses were used. The β-values were used to calculate the Euclidean distance between different samples. This distance is used to group samples in hierarchical clustering and specifies dissimilarity between samples. Principle component analysis (PCA) is an algorithm that enable simple exploration of the data by reducing the dimensionality but retaining most of the variation [171]. Principal components represent directions along which variation in the data is observed and they are widely used in genome-wide studies to visualise similarities and differences between samples [24]. PCA was performed using *prcomp* function in R. The first two principal components—PC1 and PC2—, where the variation is the data is maximal, were plotted against each other.

3.2.5. Cell composition estimates and linear model

Differentially methylated positions (DMPs) were identified by comparing uninfected and infected samples. A linear regression model was fit to the data controlling for the effects of age, sex, type 2 diabetes status, cell counts, and batch effect using the R *limma* package. The DNA methylation age of the samples were measured using Horvath's DNA methylation age calculator with the advanced analysis in blood [172]. Raw β-values (all probes on the 450K array) and a sample annotation file were uploaded to Online Age Calculator [172]. The cell counts (CD8.naive, CD8pCD28nCD45Ran, PlasmaBlast, CD4T, NK, Mono, Gran) were calculated using the methods of Houseman *et al*. [173].

3.2.6 Enrichment analyses in genomic regions

Probes annotated to regions (Illumina annotations were used) were tested for enrichment of DMPs using a two-tailed Fisher's exact test, compared to the frequency of probes in all annotated probes on the Illumina 450K array.

3.2.7 Gene Ontology analysis

Gene ontology and pathway enrichment analysis were performed using the Bioconductor package *missMethyl* [174]. The package *missMethyl* was recently published and it adopts the *goseq* method to HumanMethylation450 array. The *goseq* method was discussed in Section 2.2.4. For the enrichment analysis performed in Chapter 2, *goseq* was adopted using a custom script. In this chapter, *missMethyl* was used which employs the same methodology as the script employed in Chapter 2. The results from *missMethyl* and *goseq* analyses were in good agreement. Enriched gene ontology terms were selected based on the criteria of having a *p*-value <0.01. Redundant gene ontology terms were removed and the non-redundant terms were clustered using semantic similarity measures (the *simRel* score) using REVIGO [143].

3.2.8 Regional analysis

Enrichment analyses are more effective in identifying significant associations between DNA methylation and gene expression when applied on genomic regions rather than single positions [175]. The risk of spurious associations can be reduced when genes are identified as differentially methylated when they map to multiple DMPs located closely. A sliding window approach was employed to determine differentially methylated regions. This approach combines *p*-values of neighbouring CpG sites using Brown's method [176] and provides significance for the given genomic region (window) [175].

Each window is centred around the target CpG extending upstream and downstream by 200 bp. Brown's method [176] is an extension of Fisher's combined probability test (FCPT) [177]. FCPT combined *p*-values from a group of independent statistical tests into one test statistic. This method is successfully used for identifying differentially expressed genes [178]. In contrast, this method can be unreliable to combine differentially methylated positions, as these may be dependent. Brown's method [176] is an approximation FCPT for combining *p*-values that are not independent.

3.3 Results

3.3.1 Subject characteristics

A total of 41 subjects consisting of 28 uninfected controls and 13 patients diagnosed with septicemic melioidosis (BP) were selected for the methylation assay. The 28 uninfected controls were 13 healthy donors (H), 15 patients with type 2 diabetes (T2D), a well-documented risk factor for melioidosis. Subject characteristics are summarised in Table 3.3.1. The analyses in this chapter focused on identifying differentially methylated positions (DMPs) associated with infection.

Table 3.3.1. Characteristics of individuals included in the study.

Sample	Group	Underlying disease	Age	Gender
MK06_335	Sepsis - BP	-	55	Female
MK06_333	Sepsis - BP	T2D	39	Male
MK06_330	Sepsis - BP	T2D	43	Female
MK06_327	Sepsis - BP	COPD	68	Male
BP714	Sepsis - BP	Prev Leprosy	69	Male
BP712	Sepsis - BP	-	43	Female
BP686	Sepsis - BP	Goiter	71	Male
BP672	Sepsis - BP	T2D	72	Male
BP664	Sepsis - BP	T2D	39	Male
BP663	Sepsis - BP	T2D | HT	65	Female
BP652	Sepsis - BP	T2D | Prev TB	46	Male
BP653	Sepsis - BP	T2D	40	Female
BP683	Sepsis - BP	T2D | HT | CKD	69	Male
HD09_0553	Healthy	-	53	Female
HD09_0549	Healthy	-	51	Male
HD09_0548	Healthy	-	57	Male
HD09_0137	Healthy	-	36	Female
H369	Healthy	-	56	Female
H368	Healthy	-	60	Female
H365	Healthy	-	59	Male
H363	Healthy	-	58	Male
H359	Healthy	-	69	Male
H354	Healthy	-	64	Female
H352	Healthy	-	62	Female
H347	Healthy	-	51	Female
H346	Healthy	-	64	Female
DM08_0027	T2D	T2D	67	Male
DM08_0007	T2D	T2D	47	Male
DM08_0004	T2D	T2D	64	Female
DM08_0003	T2D	T2D	59	Male
DM526	T2D	T2D	51	Male
DM525	T2D	T2D	66	Female
DM524	T2D	T2D	39	Male
DM523	T2D	T2D	45	Female
DM522	T2D	T2D	66	Female
DM507	T2D	T2D	53	Female
DM501	T2D	T2D	66	Female
DM499	T2D	T2D	69	Male
DM339	T2D	T2D	68	Female
DM338	T2D	T2D	70	Female
DM322	T2D	T2D	75	Female

Sepsis - BP, Septicemic melioidosis; T2D, type 2 diabetes; COPD, chronic obstructive pulmonary disease; HT, hypertension; TB, tuberculosis; CKD, chronic kidney disease

3.3.2 Unsupervised analysis

To determine whether DNAm profiles of septicemic melioidosis patients were distinct from those of healthy individuals and individuals with T2D, we carried out hierarchical clustering and principal components analyses.

The set of 394048 CpG sites that passed the quality checks was used to cluster the samples according to the similarities in DNAm profiles (Figure

3.3.1a). This molecular classification was compared with the phenotype of the samples: out of the thirteen infected samples, eleven samples were clustered together. Two outlying T2D samples (DM08_0004 and DM08_003) were excluded from subsequent analyses as these samples had lower beta density compared to the other samples (Figure 3.3.1b).

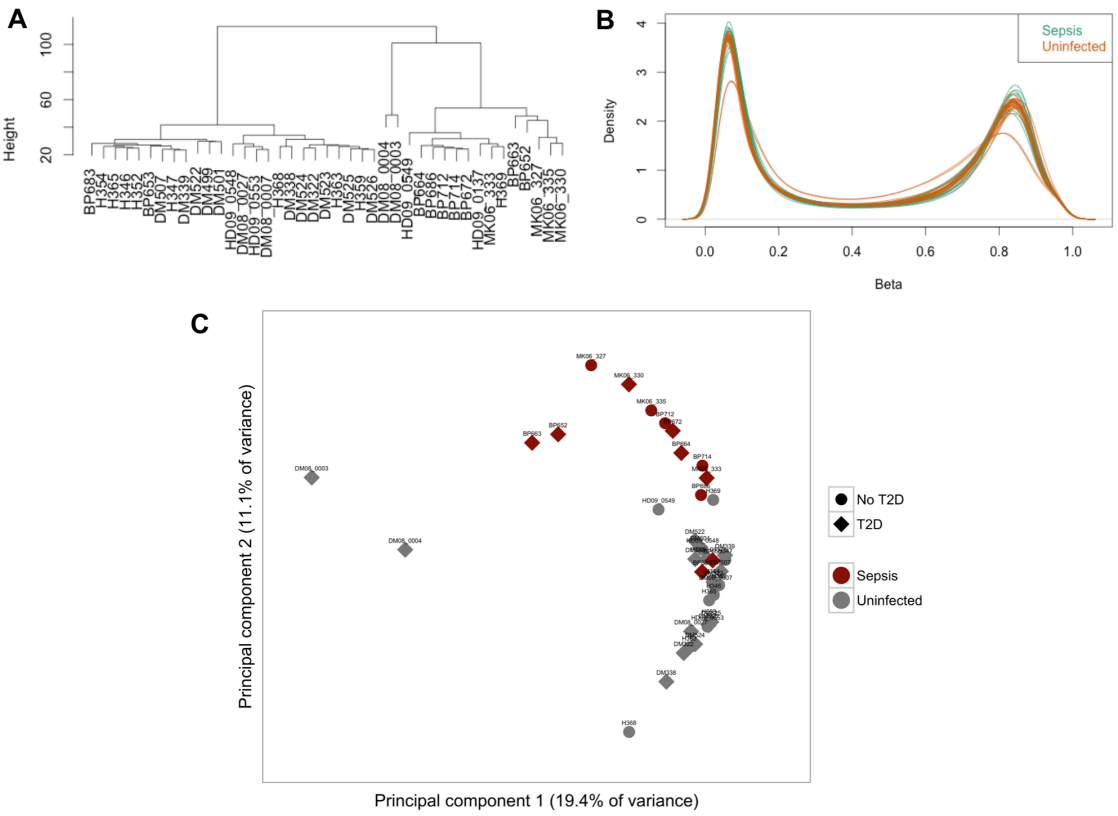


Figure 3.3.1. Clustering of the samples. (a) Hierarchical clustering. (b) Density of methylation levels. (c) Principal component analysis of DNA methylation profiles.

Furthermore, PCA was performed on methylomes on the set of 394048 CpG sites. The first two principal components explained 19.4% (PC1) and 11.1% (PC2) of the variability, and a distinct response to infected samples compared with uninfected samples was observed (Figure 3.3.1c). The molecular classification obtained through hierarchical clustering (Figure 3.3.1a)

confirmed that patients with septicemic melioidosis were separated from uninfected controls. There was no separation of healthy controls compared to T2D.

3.3.3. Identifying differentially methylated positions (DMPs)

The association of whole blood DNAm with melioidosis was examined locus-by-locus at 394048 individual CpG sites. For this, linear regression models were used, in which the dependent variable was β-value and the independent variable was melioidosis status, adjusted for age, sex, type 2 diabetes status, batch effect and cell counts. Age and cell counts were calculated using the method proposed by Horvath [172] (correlates with the chronological age, Figure 3.3.2) and Houseman [173], respectively. The distribution of *p*-values calculated using regression models for each site is shown in Figure 3.3.3.

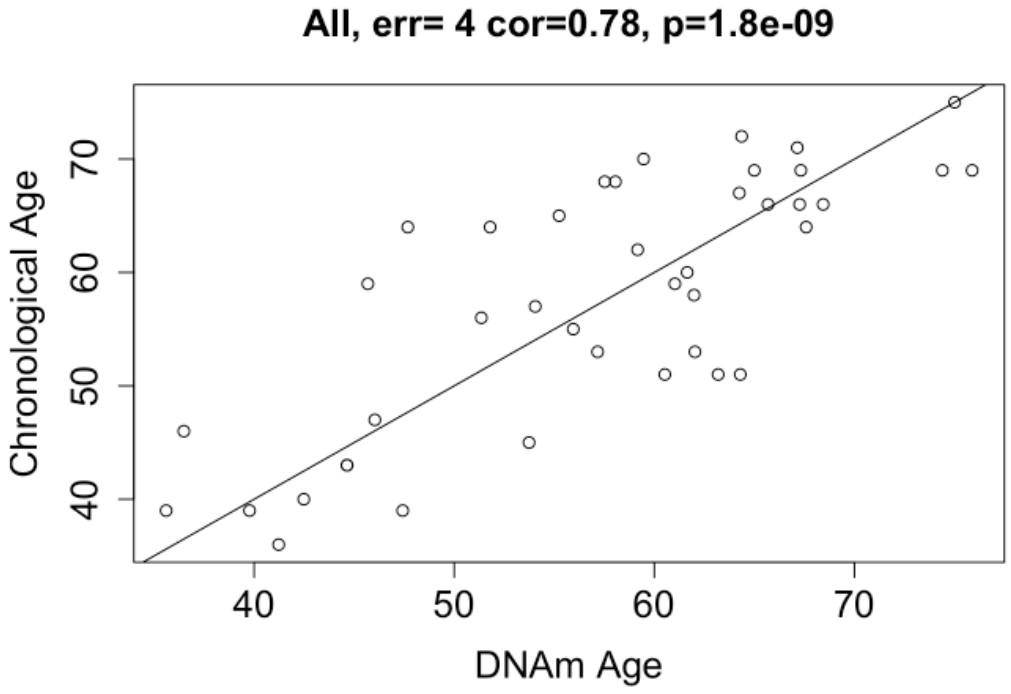


Figure 3.3.2. Chronological age versus age calculated using the method proposed by Horvath [172] **.**

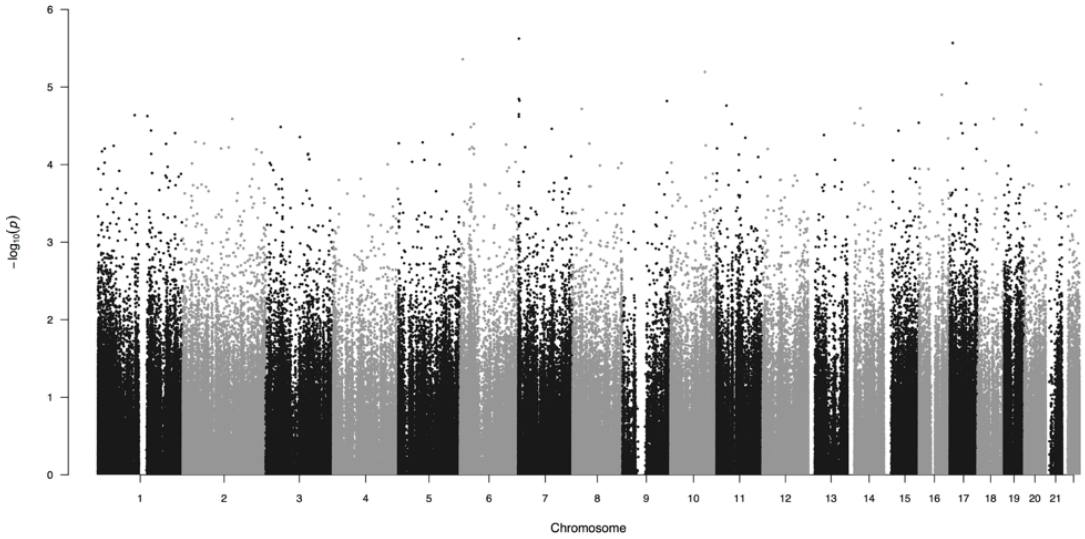


Figure 3.3.3. Manhattan plot. The distribution of *p*-values calculated using linear regression models.

Similarly, to investigate whether T2D caused DNAm changes in the host epigenome, T2D samples were compared to healthy controls using linear regression where the methylation levels were adjusted for age, sex, batch effect and cell counts.

As evidenced by the Volcano plots (Figure 3.3.4) that display –log10 *p*-values versus coefficient estimates, there were larger effect sizes when comparing melioidosis samples and controls. In contrast, no significant different methylated CpG sites were identified when comparing T2D versus healthy subjects.

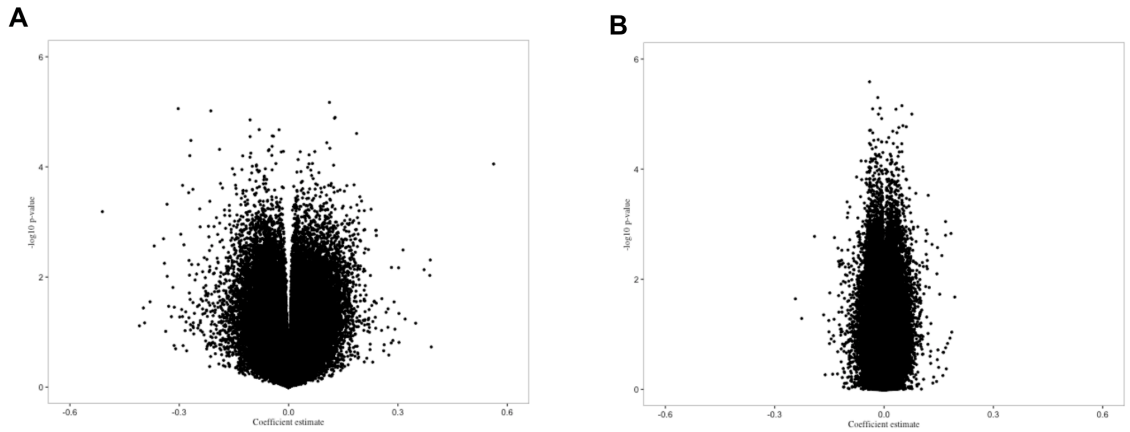


Figure 3.3.4. Volcano plots. (a) –log10 *p*-values versus coefficient estimates in melioidosis subjects. (b) –log10 *p*-values versus coefficient estimates in T2D subjects.

Next, the number of differentially methylated CpGs between melioidosis samples and controls was quantified. As an optimum level of significance for epigenome-wide association studies has yet to be established, we explored different thresholds for mean methylation level difference (Δβ) and *p*-value. As demonstrated in Figure 3.3.5, larger and more statistically significant differences are mainly losses in methylation (hypomethylation) rather than gains (hypermethylation). In fact, out of the top 100 differentially methylated positions ranked by *p*-value and absolute Δβ (Table 3.3.2), only two show hypermethylation. Mean methylation levels, Δβ, and *p*-values were uploaded to the Integrative Genomics Viewer (IGV). The MAD1L1 gene appears to have multiple DMPs within the top 100 positions (Figure 3.3.6).

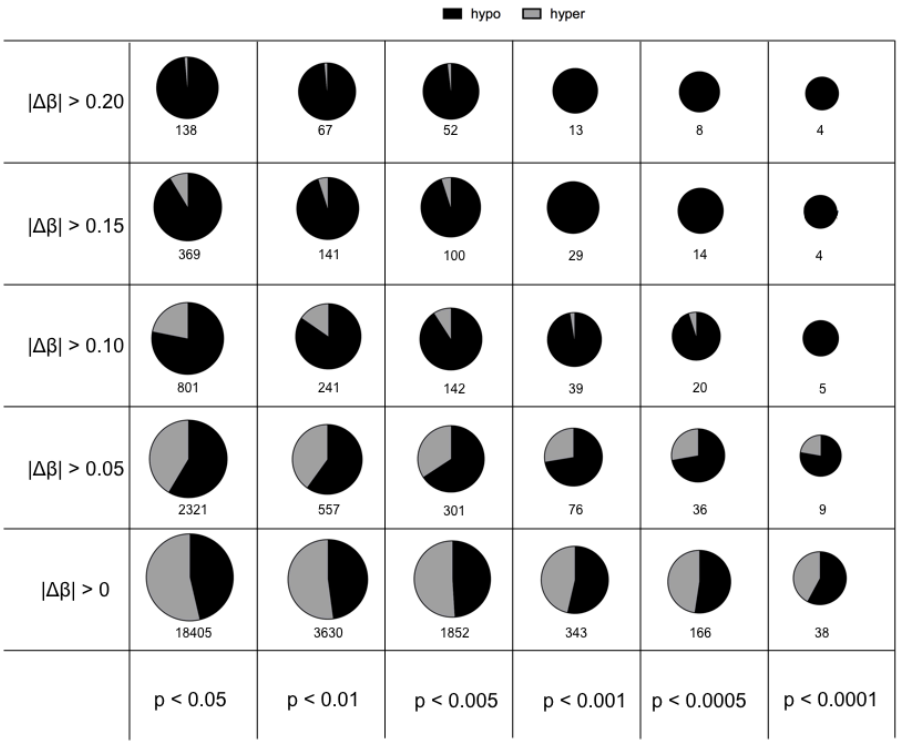


Figure 3.3.5. Number of differentially methylated positions obtained using different thresholds for mean methylation level difference (Δβ) and p-value.

Table 3.3.2. Top 100 ranked DMPs.

Probe	Chromosome	Position	Gene	Genic probe location	Relation to CpG island	Mean methylation level (β) in melioidosis samples	Mean methylation level (β) in control samples	Mean difference in methylation level (Δβ) between melioidosis and controls	P-value
cg02827175	7	1986245	MAD1L1	Body		0.3211	0.6102	-0.2891	9.61E-06
cg01952989	7	1986334	MAD1L1	Body		0.4877	0.7658	-0.2781	3.31E-05
cg16993108	7	1986301	MAD1L1	Body		0.5384	0.8151	-0.2767	6.26E-05
cg20995564	2	145172035	ZEB2	Body;Body		0.2458	0.5871	-0.3412	2.17E-04
cg01261013	21	37691747	MORC3	TSS1500	N_Shore	0.2478	0.4961	-0.2483	6.30E-05
cg05538887	2	37416883	SULT6B1	TSS1500		0.4199	0.6808	-0.2609	3.13E-04
cg04609912	16	78528146	WWOX	Body		0.4064	0.6446	-0.2382	2.46E-04
cg26204448	7	1962212	MAD1L1	Body		0.4276	0.6866	-0.2590	5.80E-04
cg20956594	13	29202889				0.1911	0.4246	-0.2335	4.01E-04
cg11826104	15	25650611	UBE3A	Body;1stExon;5'UTR		0.2966	0.5375	-0.2410	9.41E-04
cg05928849	11	122993208	ASAM	Body		0.2034	0.4332	-0.2298	8.85E-04
cg08830502	16	87588559			S_Shelf	0.3146	0.5250	-0.2104	7.02E-04
cg14195992	8	48265917	KIAA0146	Body		0.4818	0.7107	-0.2288	1.25E-03
cg11814087	5	32391135	ZFR	Body		0.2088	0.4622	-0.2534	1.57E-03
cg12644285	15	93570953	CHD2	3'UTR		0.3104	0.5421	-0.2317	1.36E-03
cg12299361	1	53796769			S_Shelf	0.1723	0.3807	-0.2085	8.72E-04
cg06526620	11	94278324	FUT4	1stExon	Island	0.2417	0.4529	-0.2112	1.03E-03
cg07110356	17	56355431	MPO	Body	Island	0.2923	0.5805	-0.2882	1.97E-03
cg15033511	17	54994023			S_Shelf	0.3591	0.5537	-0.1946	2.01E-04
cg20519581	20	48959844			N_Shore	0.1809	0.4465	-0.2656	2.09E-03
cg02969426	13	45775005	KCTD4;GTF2F2;KCTD4	5'UTR;Body;1stExon		0.3111	0.5837	-0.2725	2.25E-03
cg08122070	1	1689610	NADK	Body	N_Shelf	0.3843	0.6347	-0.2504	2.16E-03
cg12896170	6	28890069	TRIM27	Body	N_Shore	0.4639	0.6607	-0.1969	7.72E-04
cg24894783	5	10577031	ANKRD33B	Body		0.2809	0.5200	-0.2391	2.16E-03
cg00259404	6	28885568	TRIM27	Body		0.1270	0.3538	-0.2269	1.95E-03
cg05296619	13	114834683	RASA3	Body	N_Shore	0.1961	0.4182	-0.2221	1.85E-03
cg02185182	7	2185550	MAD1L1	Body	N_Shore	0.2773	0.4664	-0.1891	1.71E-04
cg26661623	17	7019262	ASGR2	TSS1500		0.2800	0.4824	-0.2024	1.13E-03
cg04958236	6	74276973				0.3597	0.5497	-0.1900	2.62E-04
cg25001190	1	61668835	NFIA	Body		0.3621	0.6029	-0.2408	2.25E-03
cg10864794	5	78203170	ARSB;ARSB	Body;Body		0.5370	0.7332	-0.1962	9.66E-04
cg07708453	1	14032034	PRDM2	1stExon;5'UTR		0.1701	0.4306	-0.2605	2.67E-03
cg03310874	7	4850260	RADIL	Body	S_Shore	0.2850	0.4792	-0.1941	9.06E-04
cg15835339	10	10828105	SFTA1P	Body		0.2214	0.4745	-0.2530	2.67E-03
cg13661827	9	111885602			S_Shelf	0.1838	0.4184	-0.2346	2.48E-03
cg26508200	12	109235071	SSH1;SSH1	Body;Body		0.2506	0.4989	-0.2483	2.70E-03
cg19982684	3	128403035			S_Shelf	0.1476	0.3345	-0.1870	4.29E-04
cg15628518	8	145025059	PLEC1	Body;TSS200	Island	0.2327	0.4612	-0.2286	2.41E-03
cg24514600	8	128805414	PVT1	TSS1500	N_Shore	0.2523	0.4988	-0.2465	2.79E-03
cg12126038	14	34966313				0.2921	0.5763	-0.2842	2.99E-03
cg19821297	19	12890029			S_Shore	0.2122	0.4398	-0.2276	2.58E-03
cg07786220	17	78683082	RPTOR	Body		0.2215	0.4029	-0.1814	1.38E-04
cg15610437	19	827821	AZU1	TSS200		0.2338	0.4219	-0.1881	7.83E-04
cg14799927	6	30175074	TRIM26	5'UTR		0.2330	0.4254	-0.1924	1.12E-03
cg23072629	2	239195537	PER2	5'UTR	N_Shore	0.4131	0.5963	-0.1832	4.96E-04
cg02579959	15	100890963	FLJ42289	TSS1500	S_Shore	0.2546	0.4844	-0.2298	2.84E-03
cg05418105	22	50981406			N_Shore	0.2860	0.5483	-0.2622	3.37E-03
cg01822050	20	304156			N_Shore	0.2274	0.4673	-0.2399	3.19E-03
cg07737292	16	56892460	MIR138-2	Body		0.2403	0.5307	-0.2903	3.49E-03
cg06122230	20	47887219	ZNFX1	Body		0.2489	0.5043	-0.2553	3.40E-03
cg20700740	1	9339683				0.2714	0.5380	-0.2666	3.59E-03
cg00052684	6	35694245	FKBP5	5'UTR	N_Shore	0.2504	0.4256	-0.1752	7.58E-04
cg16606773	20	19955806	RIN2	Body	Island	0.3970	0.6099	-0.2129	3.32E-03
cg07276358	13	114263342	TFDP1	Body	S_Shelf	0.6516	0.8644	-0.2129	3.33E-03
cg26217402	14	74238381	C14orf43	5'UTR		0.7489	0.5453	0.2036	2.95E-03
cg01948978	1	203712327	ATP2B4	3'UTR		0.3294	0.5496	-0.2202	3.74E-03
cg25814293	7	1883251	MAD1L1	Body	N_Shore	0.3736	0.6186	-0.2450	4.45E-03
cg17594003	2	145278808	ZEB2	TSS1500	N_Shelf	0.2835	0.4830	-0.1995	3.19E-03
cg13831136	6	30070342			N_Shore	0.1744	0.3763	-0.2019	3.47E-03
cg05389236	6	36635087				0.2769	0.5029	-0.2259	4.42E-03
cg22941573	20	48761279	TMEM189	Body		0.2560	0.4922	-0.2362	4.79E-03
cg12313149	2	136763629				0.1606	0.3843	-0.2237	4.56E-03
cg04354689	16	2660830	LOC652276	Body	Island	0.4753	0.6919	-0.2166	4.42E-03
cg08783253	17	40996565	AOC2	TSS200		0.2642	0.4911	-0.2269	4.71E-03
cg04739200	6	135517046	MYB	Body		0.2502	0.4436	-0.1934	3.23E-03
cg24548817	21	45774294	TRPM2	Body		0.1808	0.4080	-0.2272	4.86E-03
cg18734877	8	27297419	PTK2B	Body		0.3554	0.5184	-0.1630	6.07E-04
cg17515347	1	159047163	AIM2	TSS1500		0.2462	0.4132	-0.1670	1.23E-03
cg24450653	1	205060293	RBBP5	Body		0.3560	0.5221	-0.1660	1.14E-03
cg23500396	11	44896984	TSPAN18	5'UTR		0.4830	0.6467	-0.1637	9.18E-04
cg20954870	5	173070254				0.2316	0.4439	-0.2123	5.09E-03
cg20782252	7	2153774	MAD1L1	Body	N_Shore	0.4803	0.6591	-0.1788	2.81E-03
cg26305174	7	100463583	SLC12A9;TRIP6	Body;TSS1500	N_Shore	0.2619	0.4420	-0.1801	2.93E-03
cg25423174	3	12236303				0.2294	0.3949	-0.1655	1.33E-03
cg08418670	19	4374567	SH3GL1	Body	S_Shelf	0.3422	0.5941	-0.2519	6.16E-03
cg26471982	8	10701153			S_Shelf	0.4166	0.5751	-0.1585	7.47E-04
cg04430204	2	219537816	STK36;RNF25	Body;TSS1500	S_Shore	0.5444	0.7101	-0.1657	1.63E-03
cg10725937	6	41130978	TREM2	TSS200		0.2008	0.3968	-0.1960	4.52E-03
cg13466002	17	4621252	ARRB2	Body		0.3054	0.5551	-0.2497	6.35E-03
cg14066298	6	30297565	TRIM39	Body	S_Shelf	0.2950	0.4899	-0.1949	4.56E-03
cg27461196	19	35630106	FXYD1	5'UTR;TSS1500	N_Shelf	0.1700	0.3343	-0.1643	1.71E-03
cg19879906	19	16392219			N_Shelf	0.2713	0.4468	-0.1755	3.01E-03
cg08764162	10	31147088	ZNF438	Body;5'UTR		0.1752	0.4074	-0.2322	6.41E-03
cg24165638	19	831456	AZU1	Body	N_Shelf	0.3064	0.4863	-0.1799	3.83E-03
cg24168413	19	35630388	FXYD1	5'UTR;TSS200	N_Shore	0.2944	0.4520	-0.1576	1.28E-03
cg07979236	1	33516461			N_Shore	0.1953	0.4036	-0.2084	6.20E-03
cg21422361	7	1884392	MAD1L1	Body	N_Shore	0.4629	0.6246	-0.1616	2.07E-03
cg19893585	8	145025064	PLEC1	Body;TSS200	Island	0.2958	0.4887	-0.1929	5.30E-03
cg00008647	1	207082900	FAIM3	Body	Island	0.1519	0.3288	-0.1769	3.82E-03
cg05904013	7	128579933	IRF5	5'UTR;TSS1500	S_Shore	0.1999	0.3597	-0.1598	1.91E-03
cg09220326	3	45175421	CDCP1	Body		0.3483	0.5710	-0.2226	6.77E-03
cg14389547	19	3398778	NFIC	Body	Island	0.2575	0.4098	-0.1523	8.19E-04
cg17415111	2	86035405			N_Shelf	0.5399	0.6929	-0.1530	9.44E-04
cg01288724	5	142575122	ARHGAP26	Body;Body		0.1538	0.3249	-0.1712	3.26E-03
cg23357789	19	848026	PRTN3	3'UTR	Island	0.1960	0.3526	-0.1566	1.53E-03
cg24238409	10	93998677	CPEB3	Body	N_Shore	0.3183	0.5225	-0.2042	6.23E-03
cg25323444	7	2111060	MAD1L1	Body	S_Shelf	0.5413	0.6966	-0.1553	1.38E-03
cg17972213	1	101704898	S1PR1	Body	S_Shelf	0.7602	0.5850	0.1752	3.89E-03
cg20349598	7	1962389	MAD1L1	Body		0.5826	0.7393	-0.1567	1.68E-03
cg04454285	16	86016317				0.1333	0.3208	-0.1876	5.42E-03

Chromosomal coordinates correspond to human genome build Feb. 2009 (GRCh37/hg19).

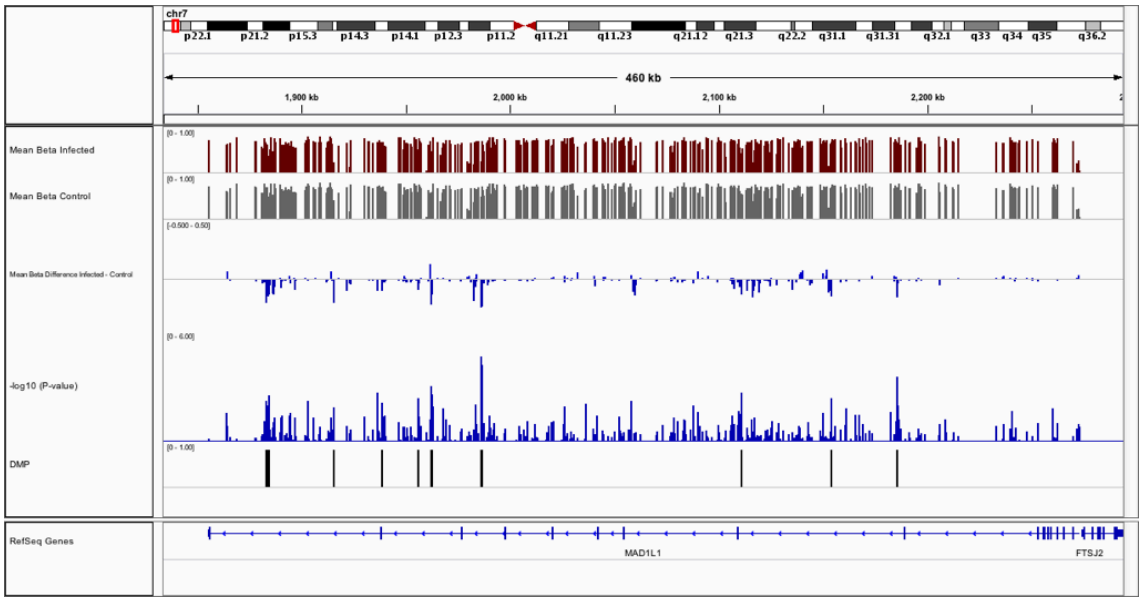


Figure 3.3.6. The MAD1L1 gene visualised using the Integrative Genomics Viewer (IGV). For each position, the mean methylation difference Δβ and corresponding *p*-value are plotted. Positions with >10% DNA methylation difference between groups with $p<0.05$ are marked in the DMP track.

Reasoning that larger differences are potentially more biologically meaningful, infection-induced DMPs (iDMPs) in melioidosis were defined as DMPs characterized by >10% DNA methylation difference between groups with $p<0.05$. We identified 801 such DMPs (78% hypomethylated, 22% hypermethylated). To better understand the observed methylomic patterns, a heatmap was generated for the DMPs (Figure 3.3.7). Differential methylation occurs at heavily, partially and sparsely methylated regions. Note that the majority of differentially methylated sites are hypomethylated.

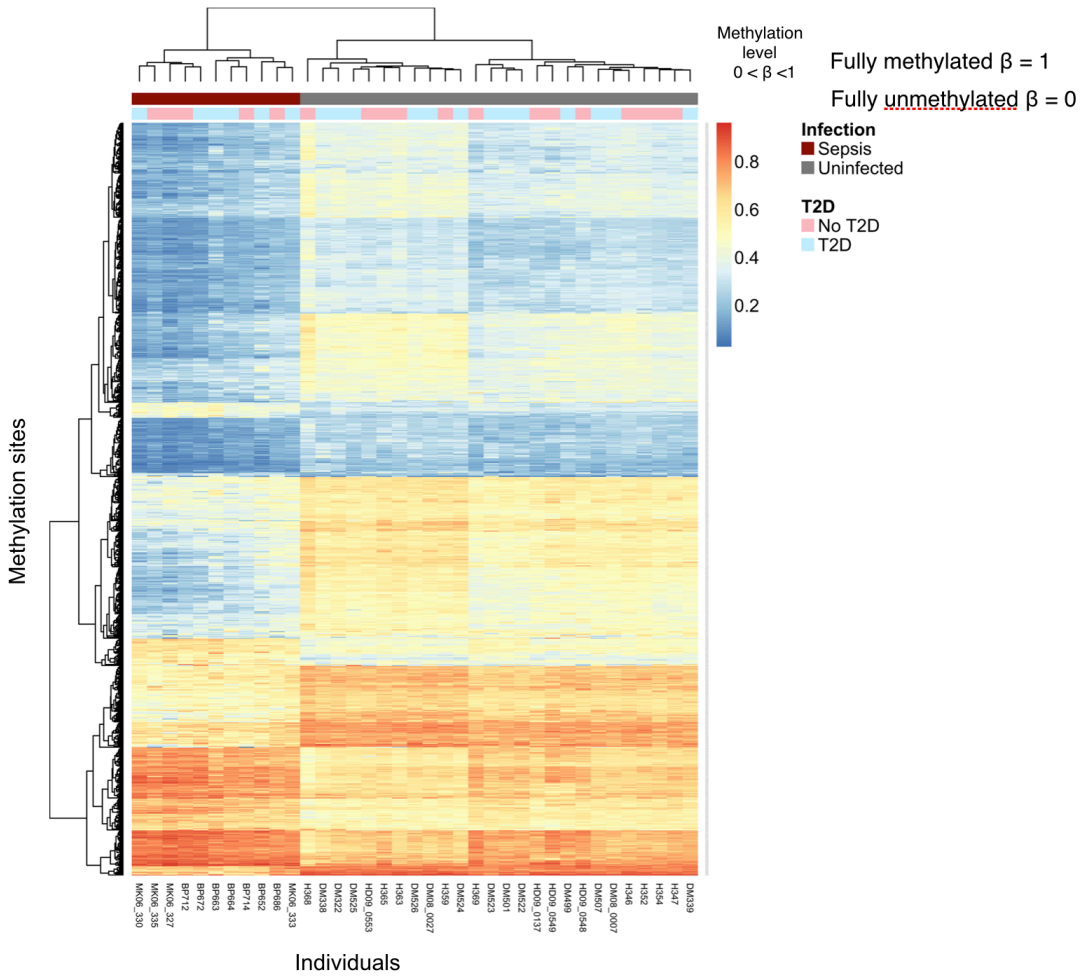


Figure 3.3.7. Heatmap of DMPs in melioidosis with >10% DNA methylation difference between groups with *p*< 0.05.

The DMPs were not uniformly distributed across the genome (Table 3.3.3). Significant over-representation of infection-induced DNA methylation was at probes located in the gene body of genes (compared to the frequency of positions in all probes on the Illumina 450K array analysed, relative enrichment (95% CI) = 1.33, p=5.11 x 10^{-6}).

Table 3.3.3. DMPs in melioidosis by genomic feature.

	Total probes	Significant probes (p<0.05 and Δβ > 0.1)	Enrichment (95% CI)	P-value	Hypermethylated	Hypomethylated
All probes	394080	801	-	-	175 (22%)	626 (78%)
GpG island feature						
Island	129166	92	0.35 (0.29 - 0.43)	1.60E-27	6 (6.5%)	86 (93.5%)
N_Shelf	18903	71	1.85 (1.43 - 2.36)	4.46E-06	21 (29.6%)	50 (70.4%)
N_Shore	52677	92	0.86 (0.68 - 1.07)	1.77E-01	15 (16.3%)	77 (83.7%)
S_Shelf	16879	60	1.75 (1.32 - 2.28)	9.63E-05	14 (23.3%)	46 (76.7%)
S_Shore	41155	72	0.86 (0.67 - 1.10)	2.47E-01	13 (18.1%)	59 (81.9%)
Not island/shore/shelf	135300	414	1.51 (1.33 - 1.70)	3.92E-11	106 (25.6%)	308 (74.4%)
Gene Feature						
TSS1500	71078	108	0.75 (0.61 - 0.91)	3.63E-03	32 (29.6%)	76 (70.4%)
TSS200	54105	56	0.51 (0.38 - 0.67)	9.94E-08	11 (19.6%)	45 (80.4%)
5' UTR	55817	120	1.06 (0.87 - 1.28)	5.48E-01	30 (25.0%)	90 (75.0%)
1st Exon	33254	45	0.67 (0.48 - 0.90)	6.82E-03	6 (13.3%)	39 (86.7%)
Gene Body	144113	390	1.33 (1.18 - 1.51)	5.11E-06	81 (20.8%)	309 (79.2%)
3' UTR	16258	36	1.09 (0.76 - 1.52)	5.94E-01	8 (22.2%)	28 (77.8%)
DNase I hypersensitivity sites	187719	536	1.41 (1.26 - 1.57)	1.78E-09	104 (19.4%)	432 (80.6%)
Transcription factor binding sites	191157	438	1.13 (1.00 - 1.27)	4.53E-02	88 (20.1%)	350 (79.9%)

To explore the biological significance of the genes mapping to the set of iDMPs, iDMPs were annotated with their corresponding gene ontology (GO) and pathway term, and searched for over-representation of categories using the package GOseq, which weights genes based on the number of probes per gene. Enriched biological functions are summarised and visualised in Figure 3.3.8 using semantic similarity, which provides a measure of functional similarity: functionally similar GO terms appear closer in the plot. Most of the enriched functional networks are relevant to immune processes. Functions such as response to stimulus and defense response shows that the host cell is reacting to bacteria. These processes instruct responses that would initiate programmes of cell death. Pyroptosis is an inflammatory response initiated in the presence of intracellular pathogens. Thus, there is a clear antimicrobial response in the host cell. Pathogens can inhibit pyroptosis to improve their own survival. Host and the pathogen compete to regulate pyroptosis [179]. DNA methylation can be the underlying mechanism for the tight control of processes such as pyroptosis, which determines important physiological outcomes such as life or death of the host. Another way of interpreting the functional significance

of iDMPs is mapping them to pathways. Pathways are more specific compared to gene ontology terms. KEGG pathways enriched in iDMPs are listed in Table 3.3.4. Toll-like receptor signalling is part of the immune system process. Pathogen invasion is sensed by Toll-like receptors [180]. Pathways such as T-cell receptor signalling, TNF signalling regulate cytokine production and ultimately determine cell survival [181]. As there are iDMPs mapping to these pathways (Table 3.3.4) , this provides evidence for the role of DNA methylation in regulating key events during infection.

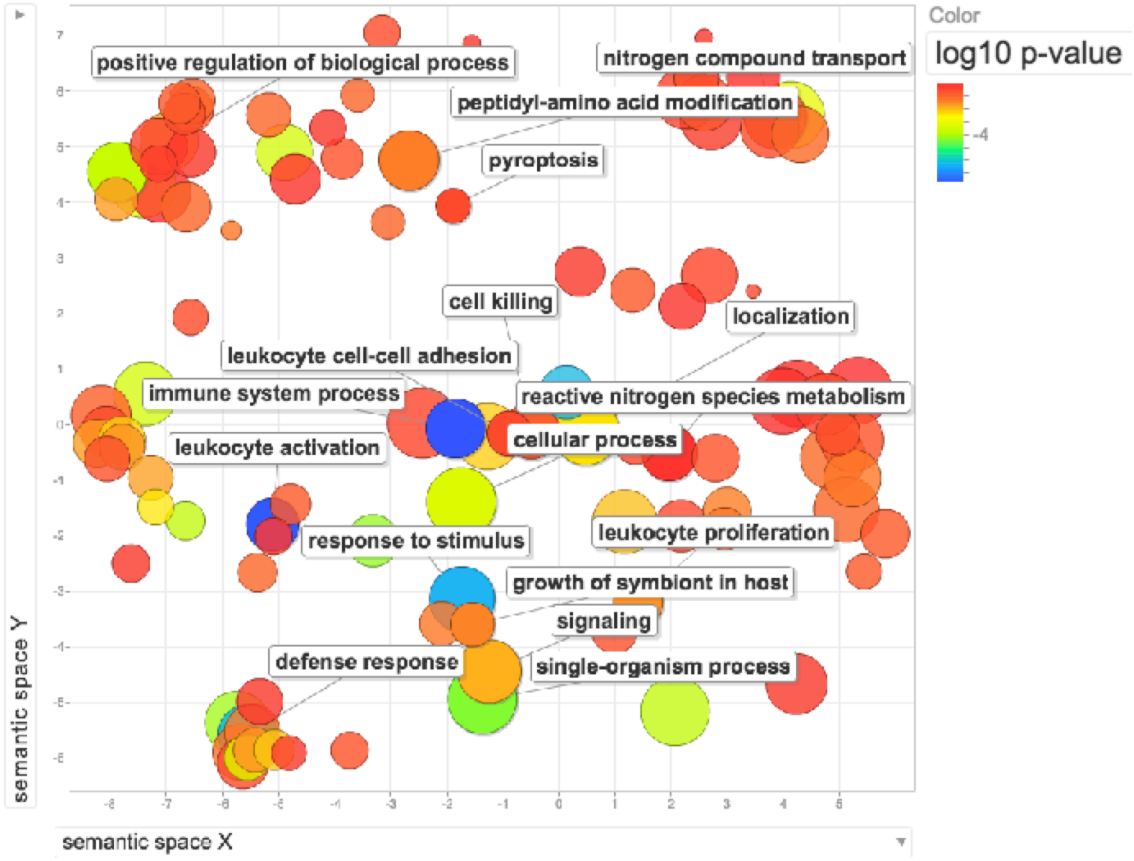


Figure 3.3.8 Gene ontology terms enriched in genes mapping to DMPs (p<0.01). The colour scale represents the *p*-values calculated using missMethyl. The non-redundant gene ontology terms are clustered using REVIGO.

Table 3.3.4. KEGG Pathways enriched (FDR<0.01) in DMPs (p<0.05 and Δβ > 0.1).

	Pathway	Number of total probes in pathway	Number of DMPs in pathway	P-value	FDR
path:hsa01100	Metabolic pathways	1160	38	1.13E-14	3.43E-12
path:hsa04650	Natural killer cell mediated cytotoxicity	108	10	4.13E-08	6.28E-06
path:hsa04152	AMPK signaling pathway	120	11	9.44E-08	7.08E-06
path:hsa04931	Insulin resistance	106	10	1.02E-07	7.08E-06
path:hsa04080	Neuroactive ligand-receptor interaction	256	13	1.16E-07	7.08E-06
path:hsa04020	Calcium signaling pathway	170	12	2.45E-07	1.24E-05
path:hsa05203	Viral carcinogenesis	195	12	3.03E-07	1.32E-05
path:hsa04710	Circadian rhythm	30	6	4.03E-07	1.53E-05
path:hsa04640	Hematopoietic cell lineage	81	7	5.09E-07	1.72E-05
path:hsa05169	Epstein-Barr virus infection	195	11	1.30E-06	3.94E-05
path:hsa04914	Progesterone-mediated oocyte maturation	85	8	1.71E-06	4.74E-05
path:hsa04922	Glucagon signaling pathway	95	8	1.92E-06	4.86E-05
path:hsa05200	Pathways in cancer	386	16	3.55E-06	8.01E-05
path:hsa04024	cAMP signaling pathway	192	11	3.69E-06	8.01E-05
path:hsa04660	T cell receptor signaling pathway	101	8	4.06E-06	8.22E-05
path:hsa04062	Chemokine signaling pathway	181	10	4.37E-06	8.31E-05
path:hsa04725	Cholinergic synapse	110	9	6.29E-06	1.12E-04
path:hsa05166	HTLV-I infection	250	12	6.80E-06	1.15E-04
path:hsa04921	Oxytocin signaling pathway	155	10	7.67E-06	1.18E-04
path:hsa04670	Leukocyte transendothelial migration	111	8	7.73E-06	1.18E-04
path:hsa04728	Dopaminergic synapse	125	9	8.51E-06	1.23E-04
path:hsa05016	Huntington's disease	180	9	1.06E-05	1.43E-04
path:hsa05145	Toxoplasmosis	112	8	1.08E-05	1.43E-04
path:hsa04060	Cytokine-cytokine receptor interaction	238	9	1.60E-05	2.03E-04
path:hsa04071	Sphingolipid signaling pathway	119	8	2.32E-05	2.77E-04
path:hsa00601	Glycosphingolipid biosynthesis - lacto and neolacto series	25	4	2.37E-05	2.77E-04
path:hsa04068	FoxO signaling pathway	127	8	2.69E-05	3.03E-04
path:hsa05164	Influenza A	152	8	2.94E-05	3.20E-04
path:hsa05152	Tuberculosis	159	8	4.49E-05	4.65E-04
path:hsa04022	cGMP-PKG signaling pathway	161	9	4.65E-05	4.65E-04
path:hsa04720	Long-term potentiation	63	6	4.74E-05	4.65E-04
path:hsa05168	Herpes simplex infection	163	8	5.51E-05	5.24E-04
path:hsa04380	Osteoclast differentiation	121	7	6.40E-05	5.89E-04
path:hsa05205	Proteoglycans in cancer	200	10	7.03E-05	6.29E-04
path:hsa05202	Transcriptional misregulation in cancer	166	9	7.94E-05	6.83E-04
path:hsa04211	Longevity regulating pathway	93	7	8.23E-05	6.83E-04
path:hsa04015	Rap1 signaling pathway	207	10	8.34E-05	6.83E-04
path:hsa04750	Inflammatory mediator regulation of TRP channels	97	7	8.60E-05	6.83E-04
path:hsa04610	Complement and coagulation cascades	76	5	8.77E-05	6.83E-04
path:hsa05340	Primary immunodeficiency	32	4	1.23E-04	9.14E-04
path:hsa04932	Non-alcoholic fatty liver disease (NAFLD)	143	7	1.23E-04	9.14E-04
path:hsa05206	MicroRNAs in cancer	275	10	1.27E-04	9.21E-04
path:hsa05332	Graft-versus-host disease	34	4	1.45E-04	1.02E-03
path:hsa05330	Allograft rejection	34	4	1.48E-04	1.02E-03
path:hsa04150	mTOR signaling pathway	150	8	1.61E-04	1.09E-03
path:hsa04144	Endocytosis	252	10	1.91E-04	1.26E-03
path:hsa04612	Antigen processing and presentation	62	5	2.36E-04	1.52E-03
path:hsa00514	Other types of O-glycan biosynthesis	30	4	2.45E-04	1.55E-03
path:hsa04910	Insulin signaling pathway	133	7	3.00E-04	1.86E-03
path:hsa04070	Phosphatidylinositol signaling system	92	6	3.36E-04	2.01E-03
path:hsa04664	Fc epsilon RI signaling pathway	66	5	3.37E-04	2.01E-03
path:hsa05161	Hepatitis B	131	7	3.55E-04	2.08E-03
path:hsa04621	NOD-like receptor signaling pathway	55	4	3.64E-04	2.08E-03
path:hsa04151	PI3K-Akt signaling pathway	315	11	3.70E-04	2.08E-03
path:hsa04925	Aldosterone synthesis and secretion	80	6	3.97E-04	2.19E-03
path:hsa04114	Oocyte meiosis	110	6	4.08E-04	2.21E-03
path:hsa05215	Prostate cancer	86	6	4.30E-04	2.29E-03
path:hsa04920	Adipocytokine signaling pathway	67	5	4.59E-04	2.40E-03
path:hsa04620	Toll-like receptor signaling pathway	89	5	4.83E-04	2.49E-03
path:hsa04940	Type I diabetes mellitus	41	4	5.73E-04	2.90E-03
path:hsa04213	Longevity regulating pathway - multiple species	63	5	7.11E-04	3.54E-03
path:hsa04261	Adrenergic signaling in cardiomyocytes	143	7	7.37E-04	3.61E-03
path:hsa04713	Circadian entrainment	94	6	8.85E-04	4.27E-03
path:hsa04915	Estrogen signaling pathway	99	6	9.67E-04	4.59E-03
path:hsa04611	Platelet activation	121	6	9.86E-04	4.61E-03
path:hsa00260	Glycine, serine and threonine metabolism	35	3	1.10E-03	5.05E-03
path:hsa05410	Hypertrophic cardiomyopathy (HCM)	80	5	1.22E-03	5.47E-03
path:hsa04210	Apoptosis	136	6	1.22E-03	5.47E-03
path:hsa04360	Axon guidance	169	8	1.40E-03	6.16E-03
path:hsa05010	Alzheimer's disease	156	6	1.56E-03	6.74E-03
path:hsa05012	Parkinson's disease	122	5	1.57E-03	6.74E-03
path:hsa04912	GnRH signaling pathway	89	5	1.73E-03	7.32E-03
path:hsa05142	Chagas disease (American trypanosomiasis)	99	5	1.83E-03	7.53E-03
path:hsa05146	Amoebiasis	95	5	1.84E-03	7.53E-03
path:hsa05222	Small cell lung cancer	82	5	1.86E-03	7.53E-03
path:hsa00350	Tyrosine metabolism	32	3	1.90E-03	7.59E-03
path:hsa04014	Ras signaling pathway	218	8	2.04E-03	8.07E-03
path:hsa05034	Alcoholism	163	6	2.29E-03	8.92E-03
path:hsa04668	TNF signaling pathway	108	5	2.57E-03	9.89E-03

3.3.4 Regional analysis

After an initial analysis of global trends and identification of DMPs, differentially methylated regions (DMRs) were identified. DMRs exhibit differences in DNA methylation on larger genomic regions rather than on single CpGs. For this, the *p*-values were combined for each of the CpGs within size ranges of few hundred bases using Brown's Method, such that neighbouring CpGs with similar differences give rise to more significant results as explained in Section 3.2.8.

3.3.5 Comparison with transcription data

To assess whether methylation expression levels change in response to infection, the transcriptome data from the same cohort were analysed (out of the 41 subjects, 28 transcriptional profiles were available (9 melioidosis, 9 healthy, 10 T2D)).

PCA analysis was performed on 7252 transcripts to explore whether transcriptional profiles exhibit a distinct response in melioidosis patients in comparison to controls. There was no separation of melioidosis samples compared to controls (Figure 3.3.9). While the transcriptomic profiles were not distinct between the groups, the methylomic profiles were distinct in melioidosis patients compared to healthy controls, as shown in Section 3.3.2. This attests to the benefits of measuring DNA methylation for detection of biomarkers for melioidosis.

To test for differential expression, melioidosis and controls were compared using the Mann-Whitney *U*-test. We did not find any statistically significant differentially expressed genes between the groups.

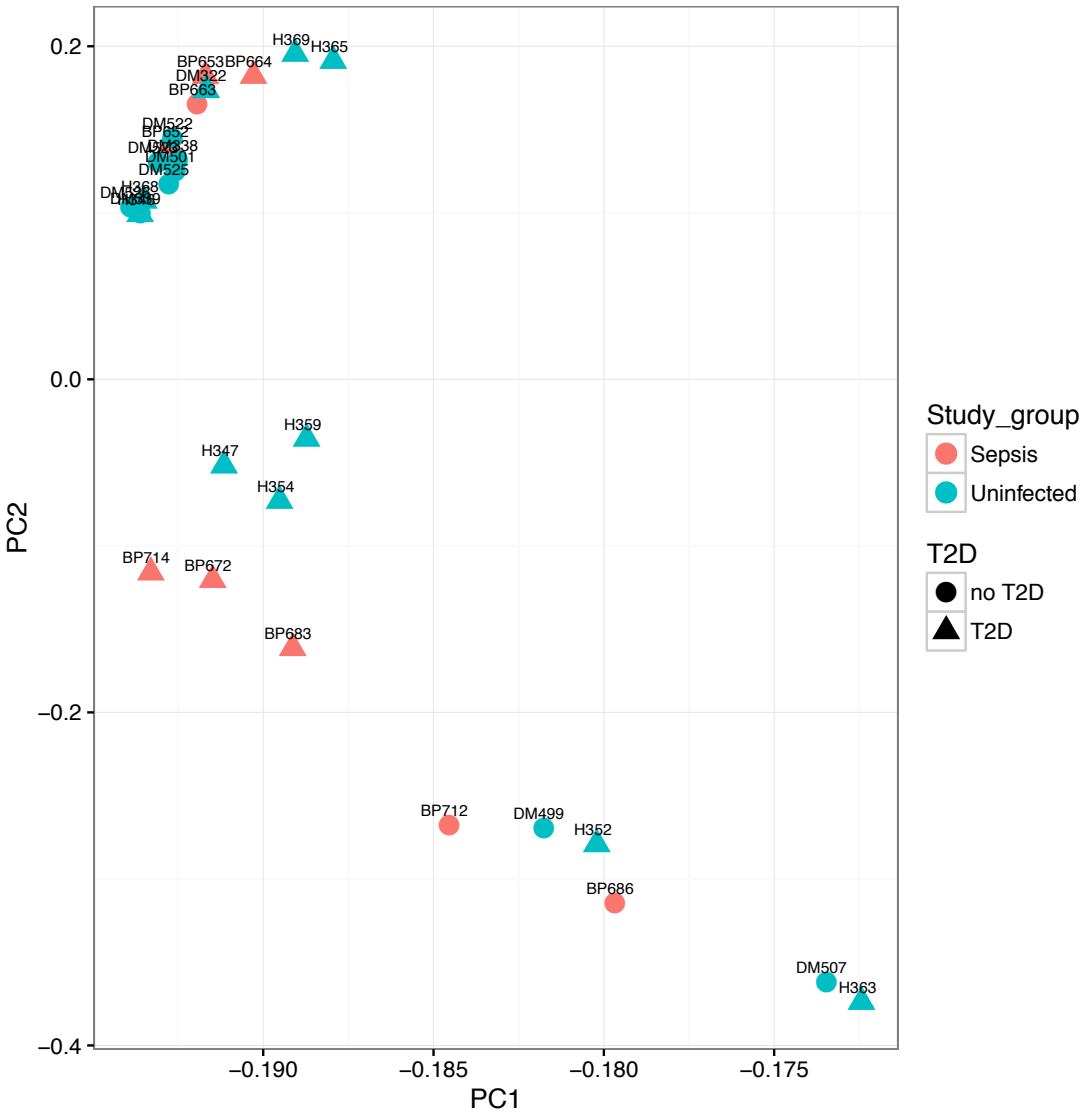


Figure 3.3.9. Principal component analysis of transcriptional profiles.

3.3.6 Comparison with other pathogens

A recent study identified epigenetic regulation in human dendritic cells before and after *in vitro* infection (at 18 hours after infection) with *Mycobacterium tuberculosis*. There were not any DMPs, which were identical in whole blood of melioidosis patients and in dendritic cells infected with *Mycobacterium tuberculosis* (Figure 3.3.10). However, when the genes to which these DMPs mapped were considered, 122 genes showed differential methylation in both studies but at different sites. These genes are likely to play

roles in pathogen-generic responses. Similar responses are expected, as melioidosis is known to mimic tuberculosis [182]. However, as the differential methylation occur in different sites, these markers can be utilised to differentiate *B. pseudomallei* infection from *M. tuberculosis*.

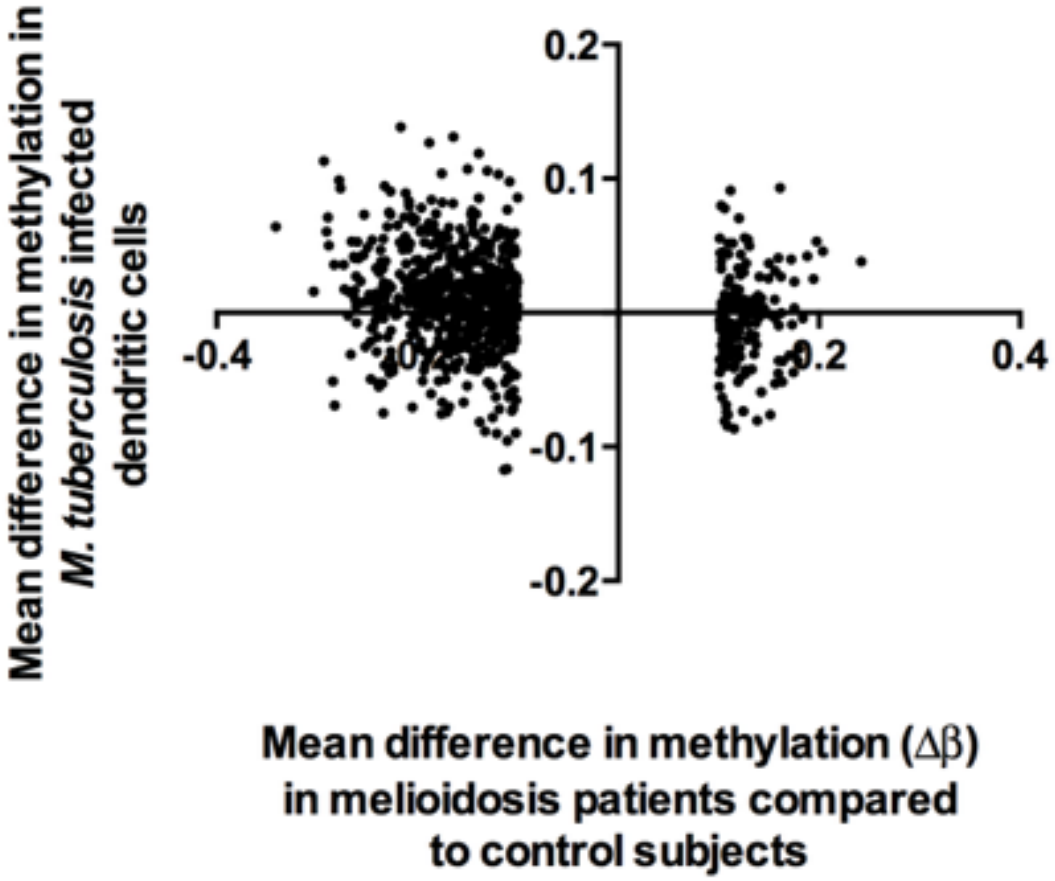


Figure 3.3.10. Comparison with publicly available DNA methylation changes during infection with *M. tuberculosis*.

3.3.7 Weighted Gene Correlation Network Analysis

Candidate biomarkers are usually identified by measuring differential regulation in disease cases compared to controls. The analyses presented in this chapter, so far, adopted this methodology and measured differential methylation. An alternative approach to identify biomarkers is to find clusters (modules) of highly correlated genes and relating these modules to external traits (i.e. disease status). Weighted Gene Correlation Network Analysis (WGCNA) is a method designed to analyse complex large gene expression data sets [183]. WGCNA identifies interacting genes by computing their correlation patterns (Pearson Correlation Coefficient). It has been successfully applied to methylomic data to identify age-related co-methylation modules [172]. In this section WGCNA analysis was applied to find co-methylated probes in melioidosis.

WGCNA R package was used to create modules of co-methylated probes; 42 modules were identified. These modules are labelled in colours. Table 3.3.6 lists the number of probes in each module. The grey module includes 130470 probes, which were not assigned to any module. Figure 3.3.11 shows the association of module eigengenes—the first principal component calculated from the DNA methylation values of all samples of that module—to the disease status. Each row corresponds to a module. Numbers in Figure 3.3.11 report the correlations of the module eigengene and traits with *p*-values in parentheses. The most significant negative correlation for melioidosis status (bp) was observed for the turquoise module and the most significant positive correlation was observed for the midnight blue module. No significant correlation was observed for diabetes status (dm) and underlying disease (dm). This shows that there are clusters of co-methylated probes that are associated

with the melioidosis status. As there are no significant associations with other diseases, modules highly correlated with melioidosis status can be utilised to find biomarkers specific to melioidosis.

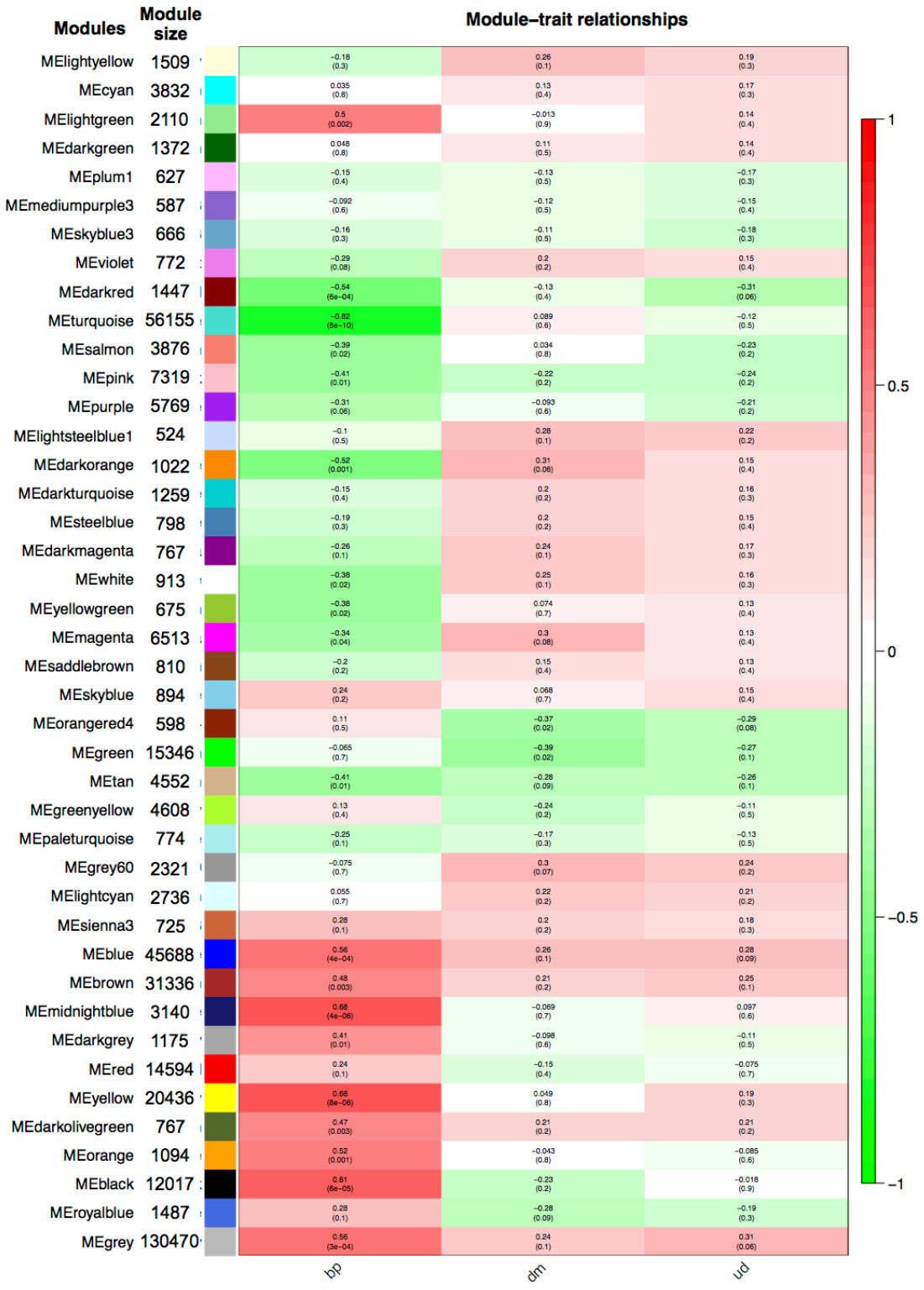
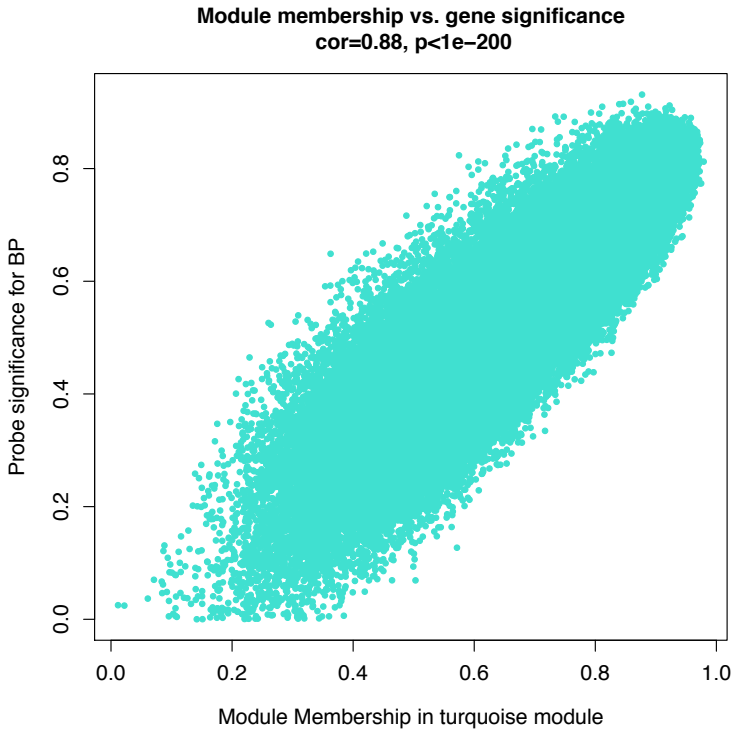


Figure 3.3.11. Module-trait relationship. bp: melioidosis; dm: diabetes T2D; ud: underlying disease. Color coded by correlation of the module eigengene (rows) and traits (columns).

In Figure 3.3.12, each point represents a probe in the given module (Figure 3.3.12a: turquoise; (b) midnightblue). For each probe the module membership is defined as the correlation between its DNA methylation value and module eigengene of the given module.

(a)



(b)

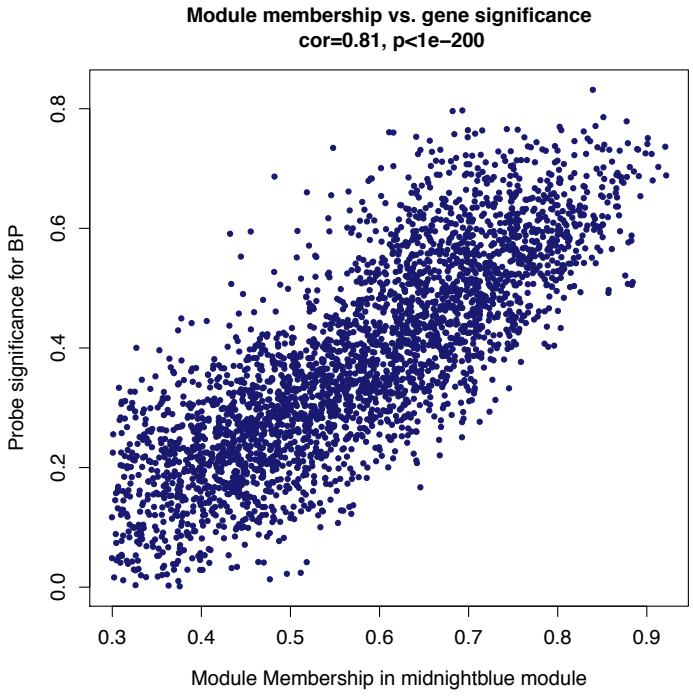


Figure 3.3.12. Each point represents a probe in the given module (Figure 3.3.12a: turquoise; (b) midnightblue). For each probe the module membership is defined as the correlation between its DNA methylation value and module eigengene of the given module.

3.4 Discussion

This chapter reports DNA methylation profiling as a promising approach to understand the pathogenesis of infection. Epigenetic signatures have been shown to remain stable long after they have been established [184,185]. In this study, molecular classification analysis showed non-separation of healthy control methylomes compared to T2D methylomes and separation of healthy controls and T2D methylomes compared to septicemic melioidosis methylomes, attesting to the existence of distinct signatures of infection. On the cohort analysed, no distinct transcriptional patterns in melioidosis patients were found. These results suggest DNAm changes related to infection can be more stably propagated through, compared to gene expression changes; thus DNAm profiles can provide reliable robust markers for classification and prognostication of infectious diseases.

The findings reported here also suggest that most of the DNAm differences in melioidosis patients are losses rather than gains in methylation. This is consistent with earlier reports of *in vitro* infection of macrophages with *B. pseudomallei*. Similar trends have been documented during infection with other pathogens [80,81]. The comparison of DMPs identified in this study with iDMPs identified in *in vitro* infection of macrophages with *B. pseudomallei* revealed no conserved DMPs. This may be due to the tissue-specific nature of epigenetic markers.

It was also found that DMPs were significantly enriched in the gene body regions. DNA methylation and gene expression are often found to be positively correlated [186].

A consensus has yet to be reached regarding optimal strategies for identifying differential methylation in disease association studies. Here, both

position and region-based approaches were adopted. Although focusing on DMRs reduces the risk of false positives, and changes across a region are more likely to control chromatin conformation and gene expression, the changes occurring at single CpG positions have not been discounted. Such changes at individual CpGs can be critical for functional variation. Moreover, although Illumina 450K covers 99% of RefSeq genes, it targets <2% of the CpG sites present in the human genome. Such technologies with sparse coverage of many genomic regions can be more suitable for interrogating single sites rather than regions. Although the *p*-values reported here do not reach the genome-wide significance estimated in many genome wide association studies, this is not surprising given the moderately small number of samples. Nonetheless, few genes were found with multiple, large consistent differences between melioidosis patients and controls.

Position and region-based approaches combined highlighted genes that are markedly differentially methylated. As listed in Table 3.3.2, Mitotic Arrest Deficient Like 1 Gene (MAD1L1) was hypomethylated at multiple positions in gene body region. Regional analysis also showed that there are multiple differentially methylated regions in the vicinity of MAD1L1 (Appendix A). MAD1L1 plays a role in cell cycle control. Deregulation of MAD1L1 can result in mitotic checkpoint failure. Hypo-methylation of MAD1L1 has been demonstrated in *M. tuberculosis* infected dendritic cells [187]. Downregulation of MAD1L1, thereby disrupting proliferation and cell cycle progression in host cells, has been suggested as an immunoevasive strategy employed by *H. pylori* [188]. *B. pseudomallei* can induce host cell cycle arrest via T3SS effector CHBP, a cell-cycle inhibitor [189]. It is possible that differential methylation of host MAD1L1 can play a role in *B. pseudomallei*-mediated cell cycle inhibition.

It was found that a range of genes associated with the immune response is differentially methylated in melioidosis patients. These genes were involved in hyper-inflammatory and immunosuppressive responses, which are associated with sepsis. Epigenetic processes are hypothesized to play a role in the fine-tuning of key biological processes. Such tight control is required for inflammatory responses, where the activation of inflammasome is crucial for pathogen recognition, while excessive inflammation can cause tissue damage. Upon infection of the host, the innate immune system engages with receptors such as Toll-like receptors (TLRs), NOD-like receptors (NLRs) and AIM2-lie receptors (ALRs) [190]. The role of NLR4 in the pathogenesis of *B. pseudomallei* is well documented [191–193]. NLR4 regulates a pro-inflammatory form of cell death called pyroptosis via the activation of caspase-1 in the early phase of *B. pseudomallei* infection [191]. Furthermore, activation of AIM2 has been demonstrated to control pyroptosis in *Francisella tularensis* [194]. This work suggests that differential methylation of NLR4, AIM2 and proinflammatory cytokine TNF can contribute to the fine control of host defense during infection.

Neutrophils are an essential part of the innate immune defense against many bacteria. Upon activation, neutrophils release neutrophil extracellular traps (NETs), which can entrap and damage bacteria [195]. It has been previously reported that *B. pseudomallei* induces formation of NETs [196] and there are abundant levels of NET-related markers in plasma from melioidosis patients [197]. Evidence was found for the differential methylation of such components, Elastase (ELANE), Myeloperoxidase (MPO) and Azurocidin 1 (AZU1). AZU1 encodes proteins with antibacterial activity contained in Azurophil granules. Azurophil granules are specialized lysosomes of the neutrophil. MPO

constitutes major component of neutrophil azurophilic granules and ELANE is involved in hydrolyzing proteins within azurophil granules. Here, it is shown that DNA methylation can underlie the regulation of bacterial containment.
Furthermore, methylation changes were also found in the vicinity of TREM-2. A recent study identifies TREM-2 as a potential treatment target for septicemic melioidosis, whereby TREM-2 deficiency leads to decreased inflammation and reduced organ damage [198].

In summary, the work presented in this chapter supports the role of DNA methylation in regulating important drivers of the pathology of infectious disease. These results demonstrate the potential clinical relevance of incorporating epigenetics information to guide the understanding of the disease. Implementing a research framework that uses studies of epigenetics would no doubt maximize our ability to develop novel treatments for complex diseases.

Whether the changes observed here are common to both septicemic melioidosis and sepsis caused by other infections awaits investigation. Moreover, it is crucial to assess the temporal changes over the course of infection and also relate to disease severity. The most robust findings may result from using longitudinally sampled cohorts of various infections distinguishing septicemic melioidosis from sepsis caused by other pathogens. This would no doubt contribute to the development of diagnostic assays.

CHAPTER 4. HOST PATHOGEN INTERACTIONS AND METABOLISM

4.1 Introduction

Common virulence factors such as capsule, flagella, type III, V and VI secretion systems of *B. pseudomallei* have been widely investigated [121,199,200], however, the large genome of *B. pseudomallei* (7.2 Mbp) accommodates a range of metabolic pathways, which still remain unstudied in terms of their role in infection. In *B. pseudomallei*, an incomplete trehalose metabolism results in lower pathogenic potential [98]. This suggests a link between metabolism and virulence.

The enzyme trehalase forms an important part of the trehalose pathway by converting trehalose into glucose. The trehalase gene has been classified as a potential virulence-associated gene in bacteria; Its presence in multiple pathogen species and absence in non-pathogenic bacteria have been demonstrated using computational comparative genomics [201]. *B. pseudomallei* is one of the pathogenic bacteria, which carry the trehalase gene (annotated as *treA*). Whole-genome tiling arrays with *B. pseudomallei* showed that *treA* is expressed in various physical, chemical, and biological conditions [202]. Vanaporn *et al.* [98] explored the role of trehalase on stress adaptation and virulence in *B. pseudomallei*. In the study of Vanaporn *et al.*, a *treA* deletion mutant (*ΔtreA*) was constructed in *B. pseudomallei* K96243 strain. The wild type *B. pseudomallei* and *B. pseudomallei ΔtreA* were compared for growth, tolerance to thermal stresses, biofilm formation, ability to survive inside macrophages, and killing in the *Galleria mellonella* and mouse infection models. The growth of the wild type and the *treA* mutant was similar in M9 minimal

medium with glucose as the carbon source. However, in M9 minimal medium with trehalose as the sole carbon source the *treA* mutant failed to grow whereas the wild type strain grew. This showed that the trehalase enzyme (TreA) is active during growth in culture medium. The *treA* mutant had increased tolerance against heat and cold stress but produced less biofilm. Moreover, the *treA* mutant had reduced ability to proliferate in macrophages and to kill *G. mellonella* and mice. Thus, mutation in trehalase gives reduced virulence even though in *treA* mutant the heat/cold stress tolerance is increased. Tolerance to stresses can differ in different environments. The heat/cold stress tolerance is measured when the bacteria grow on a medium rather than in a host cell. Within the host, the interaction of bacteria with stress-associated chemicals can be different. This interaction will depend on the nutrient availability and the influence of host signals. While current experimental findings support an important role for trehalose metabolism in *B. pseudomallei* virulence, it is very difficult to design future experiments to investigate the *B. pseudomallei* metabolism within the host environment. Many of the experimental techniques allow taking measurement in bacterial culture, however it is challenging to apply these techniques to investigate the intracellular life style of bacteria. For example, measuring concentration changes of metabolic products of bacteria growing inside a host cell is not achievable especially for Biosafety Level 3 bacteria such as *B. pseudomallei*. Therefore, *in silico* models are useful in studying the metabolic changes in highly pathogenic bacteria.

A comprehensive understanding of the complex interplay of virulence and trehalose metabolism can significantly benefit from combining experiments with computational and mathematical reasoning. Studying the interactions between the components of the biological system is key to deciphering how

organisms and biological systems work. The organisation of the network of these interactions is not always straightforward. Mathematical modelling aids the interpretation of these nonlinear processes and thus unravels the systems behaviour. In fact, mathematical models of yeast trehalose metabolism have been constructed [203,204] and used to demonstrate that yeast regulates glycolysis, the central pathway in energy metabolism, via the trehalose pathway in order to adapt to dynamic changes in the carbon source [205].

A study by van Heerden *et al.* [204] demonstrated that the trehalose pathway in yeast regulates the maintenance of functional glycolysis and facilitates the yeast to cope with stresses such as excess glucose availability. This study focused on the trehalose cycle formed by the TPS/TPP pathway (OtsA/OtsB in bacteria, Figure 4.1.1) and trehalase: In this cycle, trehalose 6-phosphate (T6P) synthase (TPS) synthesizes T6P from glucose 6-phosphate (G6P), T6P is converted to trehalose by T6P phosphatase (TPP), trehalase converts trehalose to glucose, which is phosphorylated to G6P (the first step of yeast glycolysis) completing the cycle. In particular, cells with a defect in TPS were shown to accumulate glycolytic intermediates and these cells were not able to grow on excess glucose. Smallbone *et al.* [203] also used mathematical modelling to investigate the effects of knocking out TPS and quantified the resulting changes to metabolite concentrations. The link between the TPS mutant and cellular growth has been known for a long time [206–208]. Therefore, previous mathematical modelling analyses of trehalose metabolism were limited to the TPS mutant.

This chapter tests the hypothesis that the *B. pseudomallei* trehalase mutant may be less virulent due to impaired growth. Existing mathematical models of the trehalose pathway were implemented and modified for bacterial

cells to allow the *in silico* investigation of the trehalase mutant, with the aim of evaluating whether glycolysis is affected in the trehalase mutant. Furthermore, metabolic networks of *B. pseudomallei* were analysed at the genome scale to simulate growth under different conditions and predict the effects of gene knockouts.

This chapter is organised as follows: the Methods section first discusses the modelling approaches for constructing metabolic pathways. The trehalose cycle was reviewed reaction by reaction. Knowledge of genes and the encoded enzymes of trehalose cycle were extracted from the literature and databases. Much of the kinetic information needed for constructing the trehalose cycle was obtained from articles with kinetic models of trehalose metabolism in yeast [203,209].

4.2 Methods

4.2.1 Mathematical modelling of metabolic pathways

Mathematical models of metabolism have been developed using two major modelling methodologies. Metabolic models contain either contain a few reactions described to high kinetic detail (kinetic models), or a large set of reactions with little or no kinetic information (constraint-based models) [31].

Kinetic models allow for dynamic simulations and control analysis. To build a kinetic model, rate laws need to be assigned to all reactions and the kinetic constants in each reaction must be determined. Extensive experimental data is required to characterize the mechanics of a kinetic model. For this reason, kinetic models are often available only for central pathways of well-

studied model organisms, such as *Escherichia coli* or *Saccharomyces cerevisiae* [32,33].

On the other hand, the increasing number of fully sequenced genomes allows the construction of constraint-based models for a wide range of organisms. Constraint-based models only account for the stoichiometry and directionality of reactions, which can be derived from genome annotations.

As biological processes are complex large-scale systems, complete understanding of metabolic behaviour is a challenging task. Therefore, it is crucial to exploit the existing data belonging to the organism of interest such as genome sequencing or transciptomics as well as to utilise detailed established models, which are widely described and analysed in model organisms. As the trehalose pathway is extensively studied in yeast, kinetic information on individual reactions is available. Section 4.3.1 reviews the biology and the mathematical description of reactions in the trehalose cycle sourced from yeast kinetic models [203,204]. For every reaction, in order to adapt the model to *B. pseudomallei*, the presence or absence of genes encoding for the enzymes was assessed. As kinetic parameters on *B. pseudomallei* enzymes were not available in the literature and their measurements were beyond the scope of this thesis, models with yeast or *E. coli* kinetic parameters were used to study trehalose metabolism. Sequence similarity between the *B. pseudomallei* and yeast or *E. coli* genes/proteins were calculated to evaluate whether the use of such parameters was appropriate.

4.2.2 The structure of the trehalose cycle

Kinetic models of yeast [203,204] were available in SBML format, which is a common language for computer models of biological systems. For the

purposes of building a plausible model for *B. pseudomallei* trehalose metabolism, modifications were made to the reaction kinetics and interactions of the trehalose cycle (the structure is illustrated in Figure 4.2.1) based on existing knowledge on bacterial metabolism obtained from the literature (the resultant models are referred to as modified models in this thesis). These modifications are explained in detail in Section 4.3.

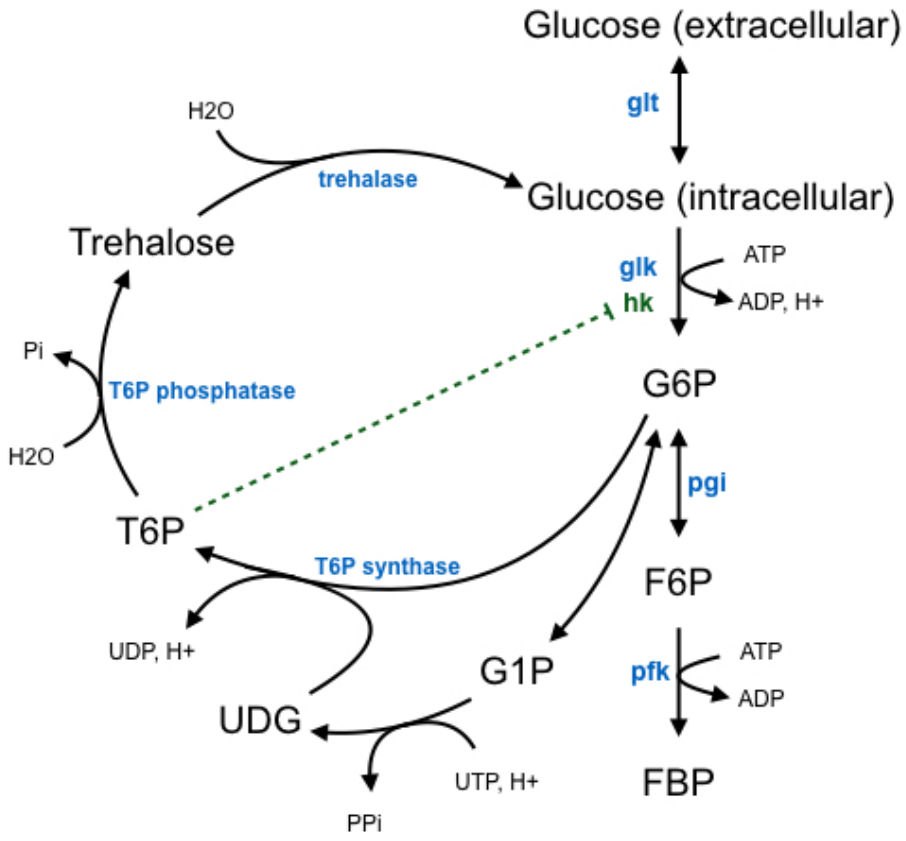


Figure 4.2.1. A structural model of the trehalose cycle. Enzymes are in blue. Evidence of allosteric inhibition of hexokinase (hk) by T6P (trehalose 6-phosphate) is found in yeast [207] (dotted line in green). G6P: glucose 6-phosphate; G1P: glucose 1-phosphate; UDG: uridine diphosphate (UDP) glucose; F6P: fructose 6-phosphate; FBP: fructose 1,6-biphosphate; glk: glucokinase; pgi: glucose 6-phosphate isomerase; pfk: phosphofructokinase. The structure is drawn using information derived from the KEGG database [26].

Sequence and structural similarity are indicative of catalytic activity and substrate specificity of proteins [210]. To reconstruct the *B. pseudomallei* trehalose pathway using kinetic information from model organisms, the similarity of *B. pseudomallei* enzymes to *E. coli* or yeast enzymes was assessed. For this, DNA and protein sequences were retrieved from NCBI [211], the *Burkholderia* Genome Database [212], the *Saccharomyces* Genome Database [213], and EcoGene 3.0 [214]. Multiple sequence comparison was carried out using the Multiple Sequence Alignment tool Clustal Omega [215]. Sequences with >30% identity were considered homologous following a common rule of thumb [216]. Furthermore, the conserved domains in *B. pseudomallei* protein sequences were identified using NCBI's Conserved Domain Database (CDD) [217]. Conserved domains provide insight into sequence, structure, and function relationships. E-values were recorded, which report the number of hits that can be expected to be seen by chance within the given location of the query sequence. E-values < 0.001 were considered significant [218].

4.2.3 Kinetic modelling of the trehalose cycle

Metabolic pathways consist of reactions, where reactants are converted into products at rates that depend on the concentrations of reactants, activators, inhibitors, and enzymes etc.

Enzymes catalyse the conversion of substrates into products. They are highly specific and can be regulated by feedback systems to provide a precise control over the rate of a reaction. Enzymatic reactions are often described using Michaelis-Menten kinetics [219], in which the reaction rate v is given by

$$v = \frac{v_{max}S}{K_M+S} \qquad \textbf{Equation 1}$$

where S is the substrate concentration, v_{max} is the maximum reaction rate, and K_M represents the concentration of substrate for which the rate of reaction is half the maximum ($v_{max}/2$).

Reaction networks can be formalized as ordinary differential equations, which represent the events of the simultaneously occurring reactions and tracks the effects on network components over time [220].

The yeast SBML models (referred to as the Smallbone and van Heerden models in this thesis) [203,204] were implemented and modified in MATLAB. The MATLAB code was formulated to facilitate modification of parameters and initial conditions of the system. The MATLAB code was able to solve the differential equations representing the system for a given combination of parameters and initial conditions, where the general form of the differential equation is,

$$\boldsymbol{x}' = \boldsymbol{S}\,\boldsymbol{f}(\boldsymbol{x}, \boldsymbol{p}), \qquad \boldsymbol{x}(\boldsymbol{0}) = \boldsymbol{x_0} \qquad \textbf{Equation 2}$$

Here, *S* is the stoichiometry matrix, metabolite concentrations are denoted by *x*, and the parameters (e.g. kinetic constants such as K_m or v_{max}) by *p*. Given the structure of the network, the parameters and initial conditions (x_0), time evolution of the variables (e.g. metabolite concentrations) can be calculated. Often the steady state values are of interest. A steady state is reached when the system is in equilibrium and the concentration of each reactant stays constant in time (*x*'=0). The computational condition used to identify a steady state was that the levels of the metabolites were constant – as assessed by an absolute tolerance value of 10^{-6} - in the last 10% of the simulation time.

4.2.4 Integration of transcriptomic data and mutants

Chieng *et al.* [126] reported expression profiles of *B. pseudomallei* genes during intracellular growth in host macrophages relative to *in vitro* growth. During intracellular infection, numerous genes involved in metabolism were down-regulated. The fold change in expression of down-regulated genes involved in the trehalose cycle is listed in Table 4.2.1.

Table 4.2.1 Trehalose metabolism related genes of *B. pseudomallei* that are down-regulated in host macrophages relative to in vitro growth [126].

Gene	Description	Log_2 Fold change (in vivo/in vitro) at the indicated time (h)			
		1	2	4	6
BPSL2410	*otsA,* alpha,alpha-trehalose-phosphate synthase	-5.90	-3.11	-3.94	-2.89
BPSL1981	*galU,* UTP--glucose-1-phosphate uridylyltransferase	-7.91	-5.48	-7.52	-2.81
BPSL1413	*pgi,* glucose-6-phosphate isomerase	-6.75	-3.33	-4.43	-2.75

In addition, genome wide transcriptome profiles of *B. pseduomallei* were available on the Pathosystems Resource Integration Center (PATRIC) Database [221]. This data included the *B. pseduomallei* transcriptome profiles obtained under exposure to 82 diverse physical, chemical, and biological conditions collected by Ooi et al. [202], and the growth transcriptome collected by Rodrigues et al. [222]. For this work, the experimental condition where the expression of *B. pseudomallei* genes in infected murine lungs were measured relative to *in vitro* growth were selected as there was a significant change in trehalase expression ($|Log_2 Ratio| >= 1$). The fold changes are listed in Table 4.2.2.

Table 4.2.2 Fold change in activity of trehalose metabolism related genes in *B. pseudomallei* in infected murine lungs relative to *in vitro* growth [202]**.**

Reaction	Locus Tag	Log_2 Fold change
Trehalase	BPSS0671	1.16
T6P phosphatase	BPSL2411	0.25
T6P synthase	BPSL2410	-0.04
Glucokinase	BPSL2614	-0.13
Glucose transport	BPSL0498	-0.01
Phosphoglucomutase	BPSL2666	0.24
UDP–glucose phosphorylase	BPSL1981	-0.22
G6P isomerase	BPSL1413	1.05

The response of *B. pseudomallei* during infection was simulated by multiplying the enzymatic activities by the factors listed in Table 4.2.1 and Table 4.2.2. Furthermore, the effects of knocking out a gene (mutants) were simulated by reducing the genes activity to 1% of its normal level. Figure 4.2.2 summarises the analysis pipeline, building the kinetic models and the incorporation of the gene activities.

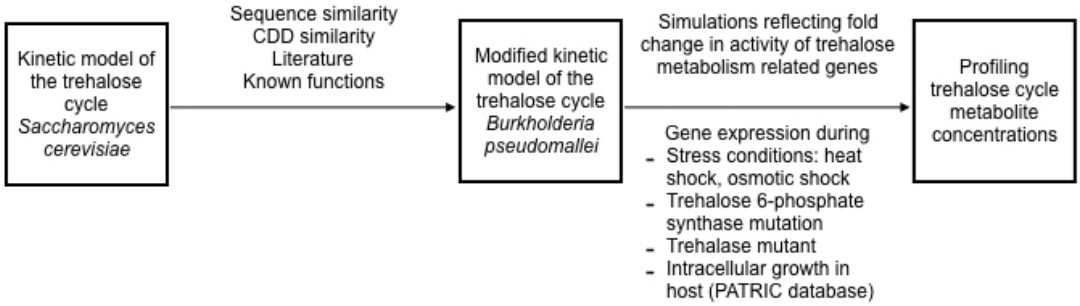


Figure 4.2.2 Analysis pipeline for the construction of kinetic model of the trehalose cycle and the integration of gene expression data

4.2.5 Sensitivity analysis

Sensitivity analysis aids the evaluation of the effect of changing a parameter on the behaviour of the model.

The sensitivities of the steady state metabolite levels with respect to 1% parameter perturbations were calculated using the following equation [34],

$$\boldsymbol{S_{i,j}} = \frac{\Delta \boldsymbol{x_i}/\boldsymbol{x_i}}{\Delta \boldsymbol{p_j}/\boldsymbol{p_j}} \approx \frac{(\boldsymbol{x_i^+} - \boldsymbol{x_i^-})/\boldsymbol{x_i}}{(\mathbf{1.01}\,\boldsymbol{p_j} - \mathbf{0.99}\,\boldsymbol{p_j})/\,\boldsymbol{p_j}} = \frac{\boldsymbol{x_i^+} - \boldsymbol{x_i^-}}{\mathbf{0.02}\,\boldsymbol{x_i}} \qquad \textbf{Equation 3}$$

Here x_i is the nominal steady state metabolite level calculated using the unperturbed parameter values, x_i^+ and x_i^- are the steady state metabolite levels that are reached when the parameter value p_j, is increased or decreased by 1%, respectively.

4.2.6 Genome-scale metabolic construction of *B. pseudomallei* and Flux Balance Analysis

The metabolic network of *B. pseudomallei* K96243 was constructed using the Model Seed [223]. Flux balance analysis (FBA) was used to assess *in silico* growth of *B. pseudomallei*. FBA for maximizing growth can be described as:

$$\text{Maximize } v_{biomass} \text{ such that } Sv = 0 \text{ and } v_{min} \leq \text{v} \leq v_{max}$$

where S is the stoichiometry matrix (rows corresponding to metabolites and the columns to reactions) and v are fluxes (conventionally in units of mmol/gDW/h) through each reaction and, $v_{min}\ and\ v_{max}$ are lower and upper bounds of the fluxes, respectively. FBA finds a solution where all the fluxes are in equilibrium, i.e the steady state $Sv = 0$ while optimising $v_{biomass}$ (the growth rate in units of

h^{-1}). FBA simulations were conducted using the COBRA Toolbox in MATLAB [224].

4.3 Results and Discussion

4.3.1 Reactions in the trehalose cycle

4.3.1.1 Trehalase

The trehalase deduced structures from bacterial, animal, plant, and fungi origin have been previously reported to be highly conserved in spite of having marked differences in amino acid sequences suggesting an evolutionary pressure to maintain the trehalose degradation function [225]. A single trehalase encoding gene (BPSS0671) has been identified in *B. pseudomallei*. The *B. pseudomallei* trehalase was similar to the *E. coli* periplasmic trehalase *treA* with 56.89% DNA sequence identity and 53.76% protein sequence identity (Table 4.3.1). In addition, *E. coli* has a cytoplasmic trehalase (TreF) and 3 other trehalase related proteins: TreB, TreC, TreR. Conserved domain search identified a neutral trehalase domain (TreA) (Table 4.3.2) supporting the presence of trehalase activity in carbohydrate transport and metabolism in *B. pseudomallei*. The *B. pseudomallei* trehalase was similar to *S. cerevisiae* neutral trehalase NTH1 (36.98% DNA sequence identity, 27.85% protein sequence identity; Table 4.3.1). Additionally, *S. cerevisiae* has an acid trehalase encoded by ATH1, and a neutral trehalase NTH2, which does not encode a trehalase activity.

Table 4.3.1. Trehalase sequence comparison. Sequence comparison was carried out using Clustal Omega. The sequence for each gene / protein is compared to *B. pseudomallei* threalase [BPSS0671 *Burkholderia pseudomallei* K96243.]

Organism	Gene	Locus Tag	Description	Percent identity in DNA sequence (%)	Percent identity in protein sequence (%)
Escherichia coli K-12 substr. MG1655	*treA*	EG11017	Periplasmic trehalase	56.89	53.76
Escherichia coli K-12 substr. MG1655	*treF*	EG12245	Cytoplasmic trehalase	55.41	48.71
Saccharomyces cerevisiae S288c	*NTH1*	YDR001C	Neutral TreHalase	36.98	27.85
Saccharomyces cerevisiae S288c	*NTH2*	YBR001C	Neutral TreHalase	36.63	27.9
Saccharomyces cerevisiae S288c	*ATH1*	YPR026W	Acid TreHalase	33.03	17.31
Escherichia coli K-12 substr. MG1655	*treB*	EG12127	Trehalose permease PTS EIIBC component	31.62	13.06
Escherichia coli K-12 substr. MG1655	*treC*	EG11402	Trehalose-6-phosphate hydrolase, osmoprotectant	31.56	11.36
Escherichia coli K-12 substr. MG1655	*treR*	EG12202	Repressor of trehalose regulon; trehalose 6-phosphate-inducible	31.34	11.56

Table 4.3.2. NCBI CD-Search engine predicted domains for TreA protein. Conserved domains identified in the protein sequences of *B. pseudomallei* trehalse.

Name	Accession	Description	Interval	E-value
Trehalase super family	cl17346	Trehalase (EC:3.2.1.28)	70-565	0.00E+00
TreA	COG1626	Neutral trehalase [Carbohydrate transport and metabolism]	72-564	0.00E+00

4.3.1.2 T6P phosphatase and T6P synthase

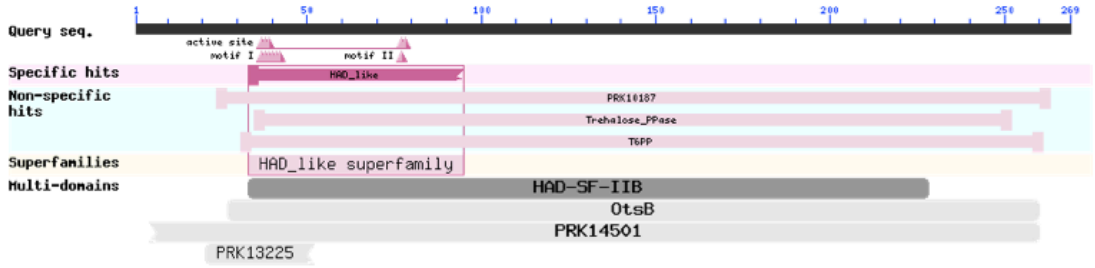
Trehalose biosynthesis through the trehalose-6-phosphate synthase (TPS) - trehalose-phosphatase (TPP) route is the most widely distributed pathway across kingdoms [106]. While bacteria have five biosynthetic pathways known for trehalose, yeast has only two, with the TPS/TPP pathway being the main pathway. In this two-step process, glucose 6-phosphate and UDP-glucose are converted by TPS (encoded by *TPS1* in yeast, encoded by *otsA* in bacteria) into trehalose 6-phosphate, which is then converted with water into trehalose

and phosphate by TPP (encoded by *TPS2* in yeast, encoded by *otsB* in bacteria). In yeast, the TPS/TPP pathway received wide attention as the intermediate metabolite trehalose 6-phosphate can affect glycolysis by inhibiting hexokinase, which is the enzyme involved in the first step of glycolysis [226]. This is reviewed next, in Section 4.3.1.3. Sequence comparison showed that *B. pseudomallei* trehalose-phosphatase is similar to *E. coli* and *S. cerevisiae* trehalose-phosphatase (Table 4.3.3). CD-Search confirmed the presence of Trehalose-6-phosphatase (OtsB) activity in *B. pseudomallei* trehalose-phosphatase protein (Table 4.3.1). Two genes were identified in *B. pseudomallei* for trehalose-phosphate synthase: BPSL2410 and BPSL1044. These two *B. pseudomallei* trehalose-phosphate synthases exhibit high homology (70.10% DNA sequence identity and 61.72% protein identity) and were similar to *E. coli* OtsA and *S. cerevisiae* TPS1, respectively (Table 4.3.5 and Table 4.3.6). CD-Search was conducted on the protein sequence encoded by BPSL2410 and the results showed OtsA activity (Table 4.3.7)

Table 4.3.3. Trehalose-phosphatase sequence comparison (otsB).
Sequence comparison was carried out using Clustal Omega. The sequence for each gene / protein is compared to *B. pseudomallei* trehalose-phosphatase [BPSL2411 *otsB* trehalose-phosphatase cytoplasmatic]

Organism	Gene	Locus Tag	Description	Percent identity in DNA sequence (%)	Percent identity in protein sequence (%)
Escherichia coli K-12 substr. MG1655	*otsB*	EG11752	Trehalose phosphate phosphatase	49.20	42.11
Saccharomyces cerevisiae S288c	*tps2*	YDR074W	Trehalose-phosphatase	37.08	22.64

Table 4.3.4. NCBI CD-Search engine predicted domains for OtsB protein. Conserved domains identified in the protein sequences of *B. pseudomallei* trehalose-phosphatase.



Name	Accession	Description	Interval	E-value
PRK10187	PRK10187	trehalose-6-phosphate phosphatase	24-263	1.48E-80
Trehalose_PPase	pfam02358	This family consist of trehalose-phosphatases EC:3.1.3.12	35-252	4.60E-55
T6PP	TIGR00685	trehalose-phosphatase	31-261	7.35E-53
HAD_like	cd01427	Haloacid dehalogenase-like hydrolases	33-95	1.42E-06
HAD-SF-IIB	TIGR01484	HAD-superfamily hydrolase, subfamily IIB	33-228	4.81E-24
OtsB	COG1877	Trehalose-6-phosphatase [Carbohydrate transport and metabolism]	27-260	1.67E-70
PRK14501	PRK14501	putative bifunctional trehalose-6-phosphate synthase/HAD hydrolase subfamily IIB	5-260	1.74E-40
PRK13225	PRK13225	phosphoglycolate phosphatase	21-52	7.60E-03

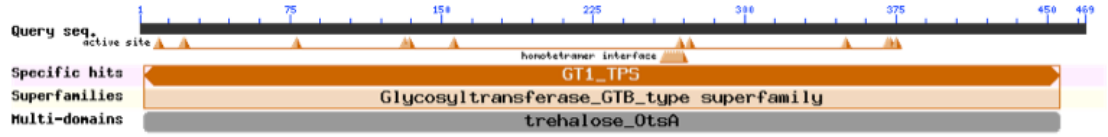
Table 4.3.5. Trehalose-phosphatase sequence comparison (otsA). Sequence comparison was carried out using Clustal Omega. The sequence for each gene / protein is compared to *B. pseudomallei* threalase [BPSL2410 alpha,alpha-trehalose-phosphatase synthase Cytoplasmatic]

Organism	Gene	Locus Tag	Description	Percent identity in DNA sequence (%)	Percent identity in protein sequence (%)
Burkholderia pseudomallei K96243	*otsA* (UDP-forming)	BPSL1044	alpha,alpha-trehalose-phosphate synthase	70.10	61.72
Escherichia coli K-12 substr. MG1655	*otsA*	EG12245	Trehalose phosphate synthase	55.34	51.72
Saccharomyces cerevisiae S288c	*TPS1*	YDR001C	Trehalose 6-phosphate synthase	43.15	31.68

Table 4.3.6. Trehalose-phosphatase sequence comparison (otsA UDP-forming). Sequence comparison was carried out using Clustal Omega. The sequence for each gene / protein is compared to *B. pseudomallei* threalase [BPSL1044 alpha,alpha-trehalose-phosphatase synthase Cytoplasmatic]

Organism	Gene	Locus Tag	Description	Percent identity in DNA sequence (%)	Percent identity in protein sequence (%)
Burkholderia pseudomallei K96243	*otsA*	BPSL2410	alpha,alpha-trehalose-phosphate synthase	70.10	61.72
Escherichia coli K-12 substr. MG1655	*otsA*	EG12245	Trehalose phosphate synthase	54.66	50.85
Saccharomyces cerevisiae S288c	*TPS1*	YDR001C	Trehalose 6-phosphate synthase	44.32	31.83

Table 4.3.7. NCBI CD-Search engine predicted domains for OtsA protein. Conserved domains identified in the protein sequences of *B. pseudomallei* trehalose-phosphate synthase.



Name	Accession	Description	Interval	E-value
GT1_TPS	cd03788	Trehalose-6-Phosphate Synthase (TPS)	3-456	0.00E+00
trehalose_OtsA	TIGR02400	alpha,alpha-trehalose-phosphate synthase [UDP-forming]	3-456	0.00E+00

4.3.1.3 Glucokinase

The first step in metabolising intracellular glucose involves phosphorylation of glucose to form glucose 6-phosphate. In yeast, this reaction can be catalysed by three enzymes: hexokinases (HXK1 and HXK2), and glucokinase (GLK1) [227]. The trehalose pathway can regulate yeast glycolysis via T6P acting as a competitive inhibitor of hexokinases to glucose; the regulation is mainly through the inhibition of hexokinase II, while no inhibition is observed upon glucokinase [207].

In most bacteria, glucose is transported by the phosphotransferase system (PTS) as glucose-6-phosphate [228]. Therefore, glucokinase is not used in the utilisation of extracellular glucose. For this reason, the glucokinase reaction is not included in bacterial kinetic models, which focus on the central metabolism such as glycolysis. However, in metabolism of disaccharides such as lactose, maltose, or trehalose, glucokinase has an important role of phosphorylating the glucose formed inside the cell.

Glucokinase and hexokinases differ in their kinetic properties. Glucokinase has a lower affinity (high Km) for glucose than the other hexokinases and therefore is effective only when glucose is abundant [229]. Glucokinase is not inhibited by its product, glucose-6-phosphate; the rate of

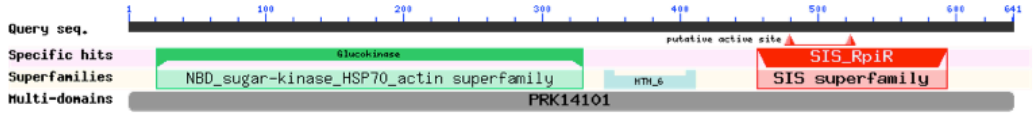
reaction is driven by the supply of glucose, not by the demand for end products [230].

Meyer *et al.* reported the kinetic properties of glucokinase in *E. coli* [230]. In yeast, hexokinase is modelled to allow for the competitive inhibition of glucose by T6P. Here, in the modified model, this step is modelled following the kinetics of glucokinase [230].

Table 4.3.8. Glucokinase sequence comparison. Sequence comparison was carried out using Clustal Omega. The sequence for each gene / protein is compared to *B. pseudomallei* threalase [Glk BPSL2614 bifunctional glucokinase/RpiR family transcriptional regulator Cytoplasmatic]

Organism	Gene	Locus Tag	Description	Percent identity in DNA sequence (%)	Percent identity in protein sequence (%)
Escherichia coli K-12 substr. MG1655	*glk*	EG12957	Glucokinase	51.98	45.77
Saccharomyces cerevisiae S288c	*GLK1*	YCL040W	GlucoKinase	41.41	19.69

Table 4.3.9. NCBI CD-Search engine predicted domains for OtsA protein. Conserved domains identified in the protein sequences of *B. pseudomallei* glucokinase.



Name	Accession	Description	Interval	E-value
Glucokinase	pfam02685	Glucokinase; This is a family of glucokinases or glucose kinases EC:2.7.1.2.	21-329	3.36E-140
SIS_RpiR	cd05013	RpiR-like protein.	456-593	4.32E-40
HTH_6 super family	cl15975	Helix-turn-helix domain, rpiR family	345-411	9.35E-17
PRK14101	PRK14101	bifunctional glucokinase/RpiR family transcriptional regulator	1-641	0.00E+00

Table 4.4.10 summarises the rate laws used in the kinetic model of the trehalose cycle. These rate laws have been adapted from the Smallbone Model [203]. Smallbone *et al.* retrieved much of the kinetic information from a series of publications [231–234]. In the Smallbone Model, the conversion of glucose to glucose 6-phosphate is modelled with hexokinase kinetics allowing for the competitive inhibition of glucose by trehalose 6-phosphate. In this chapter, in the Modified Smallbone Model, the hexokinase reaction is replaced with the glucokinase reaction for the reasons explained in Section 4.3.1.3. The resulting rate law is presented in Table 4.3.10.

Table 4.3.10. Reaction laws for each step of the trehalosecycle for the Modified Smallbone Model.

Enzyme	Reaction	Rate law	Reference
Trehalase	Trehalose + water → 2 glucose	$\frac{v}{Vmax} = \frac{\frac{[TRE]}{K_{TRE}}}{1+\frac{[TRE]}{K_{TRE}}}$	[65]
T6P phosphatase	Trehalose 6-phosphate + water → trehalose + phosphate	$\frac{v}{Vmax} = \frac{\frac{[T6P]}{K_{T6P}}}{1+\frac{[T6P]}{K_{T6P}}}$	[64]
T6P synthase	Glucose 6-phosphate + UDP glucose → trehalose 6-phosphate + UDP + H^+	$\frac{v}{Vmax} = \frac{\frac{[G6P]}{K_{G6P}}\frac{[UDPG]}{K_{UDPG}}}{\left(1+\frac{[G6P]}{K_{G6P}}\right)\left(1+\frac{[UDPG]}{K_{UDPG}}\right)}$	[62]
Glucokinase	Glucose + ATP → glucose 6-phosphate + ADP + H^+	$\frac{v}{Vmax} = \frac{\frac{[GLC]}{K_{GLC}}\frac{[\square TP]}{K_{ATP}}}{\left(1+\frac{[GLC]}{K_{GLC}}\right)\left(1+\frac{[ATP]}{K_{ATP}}\right)}$	
Phosphoglucomutase	Glucose 6-phosphate ⇌ glucose 1-phosphate	$\frac{v}{Vmax} = \frac{\frac{1}{K_{G6P}}\left([G6P]-\frac{[G1P]}{K_{eq}}\right)}{\left(1+\frac{[G6P]}{K_{G6P}}+\frac{[G1P]}{K_{G1P}}\right)}$	[29]
UDP glucose pyrophosphorylase	G1P -> UDP-glucose Glucose 1-phosphate + UTP + H^+ → UDP glucose + diphosphate	$\frac{v}{Vmax} = \frac{\frac{[UTP]}{K_{UTP}}\frac{[G1P]}{K_{G1P}}}{\frac{K_{UTP}}{K_{UTP}}+\frac{[UTP]}{K_{UTP}}+\frac{[G1P]}{K_{G1P}}+\frac{[UTP]}{K_{UT\square}}\frac{[G1P]}{K_{G1P}}+\frac{K_{UTP}}{K_{UTP}}\frac{[UDPG]}{K_{UDPG}}+\frac{[G1P]}{K_{G1P}}\frac{[UDPG]}{K_{UDPG}}}$	[65]

4.3.2 Response of yeast kinetic model (Smallbone Model) to stress conditions and mutation

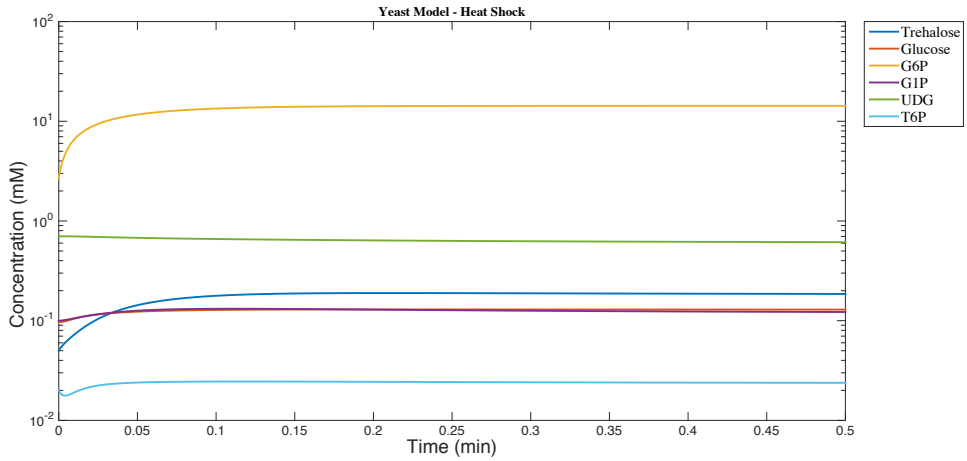
The kinetic model of the trehalose cycle in yeast developed by Smallbone *et al.* [203] investigated the effects of heat shock, osmotic shock, and the *TPS1* mutation on the concentrations of metabolites directly involved in trehalose metabolism. In this section, firstly, the simulation results demonstrated by Smallbone *et al.* were verified. This allowed testing of the model and to avoid any errors in translating the SBML file into MATLAB code. The steady state concentrations of the metabolites in response to heat shock and *TPS1* mutation generated by Smallbone *et al.* [203], are shown in columns 2 and 4 of Table 4.3.11. These values were successfully reproduced (column 3 and 5 of Table 4.3.11) by simulating the response of the metabolites to heat shock and *TPS1* mutation (Figure 4.3.1 A and Figure 4.3.1 C). The system takes a longer time to reach the steady state in the *TPS1* mutant compared to heat and osmotic shock (Figure 4.3.1). This shows that the system can tolerate heat and osmotic shock using the trehalose cycle. However, when the trehalose cycle is impaired (*TPS1* mutant), the system is imbalanced for a longer time. This can have adverse effects on the yeast cell.

In addition, the trehalase mutant was simulated in the yeast model (Figure 4.3.2) by reducing the trehalase activity to 1% of its normal level; in this case trehalose accumulates and the system cannot reach equilibrium (simulations are carried out for a long time (10^50 min) to show that trehalose levels keep building up to biologically implausible levels ~10^30 mM; Figure 4.3.2B).

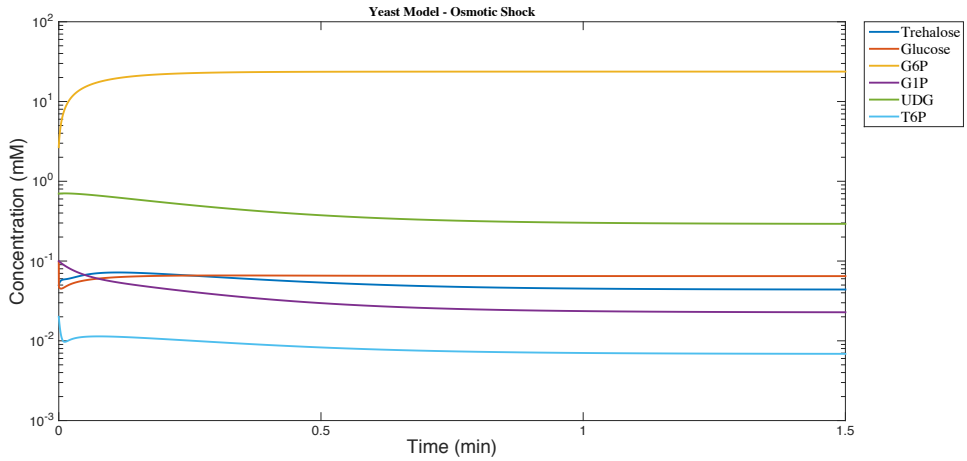
Table 4.3.11. Smallbone Model concentrations in response to heat shock, osmotic shock and, *TPS1* mutation and trehalase mutation.

Metabolite	Heat shock (as presented by Smallbone *et al.*)	Heat shock (simulated using the MATLAB code)	TPS1 mutant (as presented by Smallbone et al.)	TPS1 mutant (simulated using the MATLAB code)	Osmotic shock	Trehalase mutant (simulation time - 10^50 min)
Glucose	0.121	0.1246	-0.146	-0.142	-0.1748	0.0024
Glucose 6-phosphate	0.728	0.7278	0.00213	0.0021	0.9478	-0.00022037
Glucose 1-phosphate	0.0787	0.0787	0.642	0.6418	-0.6442	2.11E-05
UDP glucose	-0.0636	-0.0636	2.05	2.0482	-0.3805	1.44E-04
Trehalose 6-phosphate	0.0718	0.0718	-1.67	-1.6651	-0.4658	-7.53E-05
Trehalose	0.565	0.5646	-1.66	-1.6555	-0.0576	50.29

A



B



C

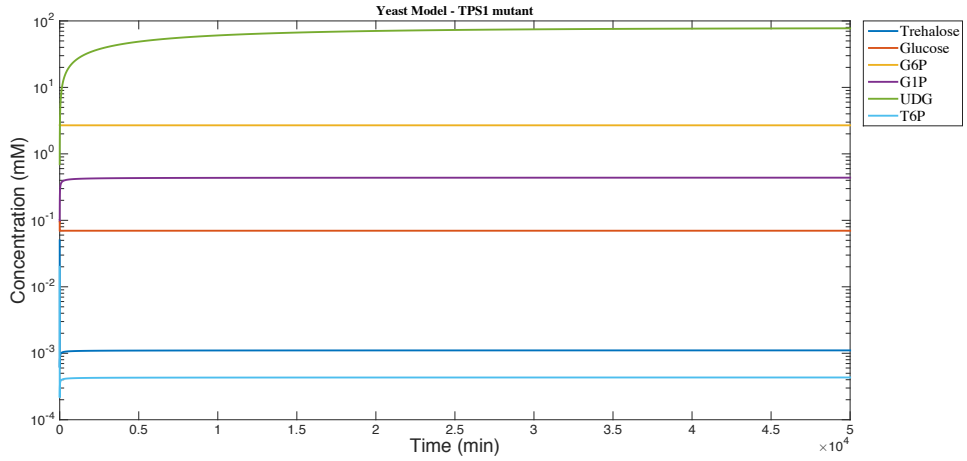
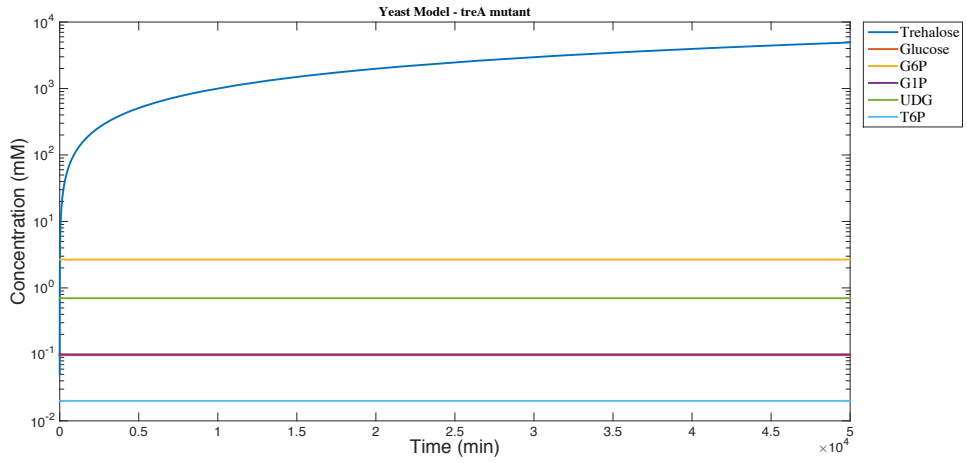


Figure 4.3.1. Smallbone Model concentrations in response to (A) heat shock, (B) osmotic shock, (C) TPS1 mutation. Initial concentrations and parameters were taken from Smallbone et al. [203]. Changes in enzymatic

activities induced by heat and osmotic shock were taken from Smallbone et al. [203]. TPS1 mutation was simulated by reducing TPS1 activity to 1% of the normal level.

A



B

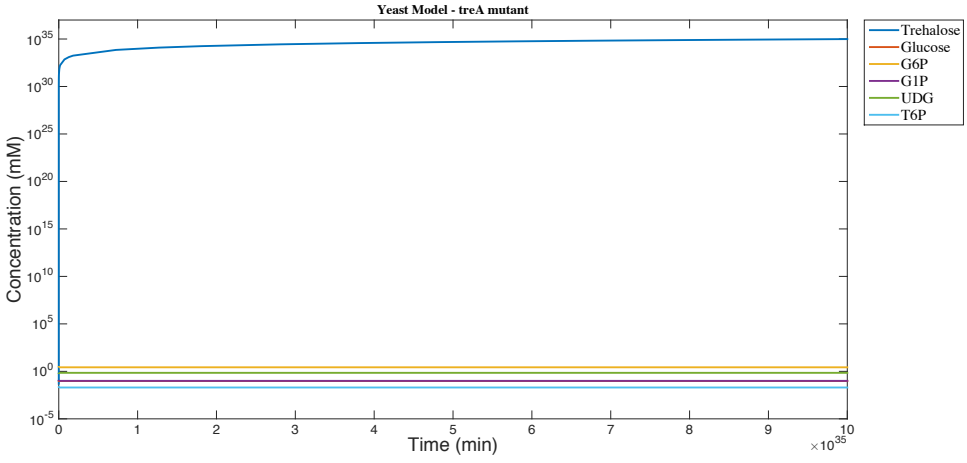


Figure 4.3.2. Trehalase mutation in yeast. Smallbone Model concentrations in response to trehalase mutation (Trehalase activity is reduced to 1% of its normal level). Initial concentrations and parameters were taken from Smallbone et al. [203]. (A) simulation time $5x10^4$ (B) simulation time 10^{35}.

4.3.3 Response of the Smallbone Modified Model to stress conditions and mutation

The model was modified by replacing the hexokinase kinetics with glucokinase kinetics as explained in Section 4.3.1.3 (the new model is referred to as Smallbone Modified Model). As this has introduced a change to the system, a simulation was carried out to restore the system into a new equilibrium (Figure 4.3.3). The steady state values that were reached in this simulation were accepted as initial conditions for the subsequent simulations when the Smallbone Modified Model was used.

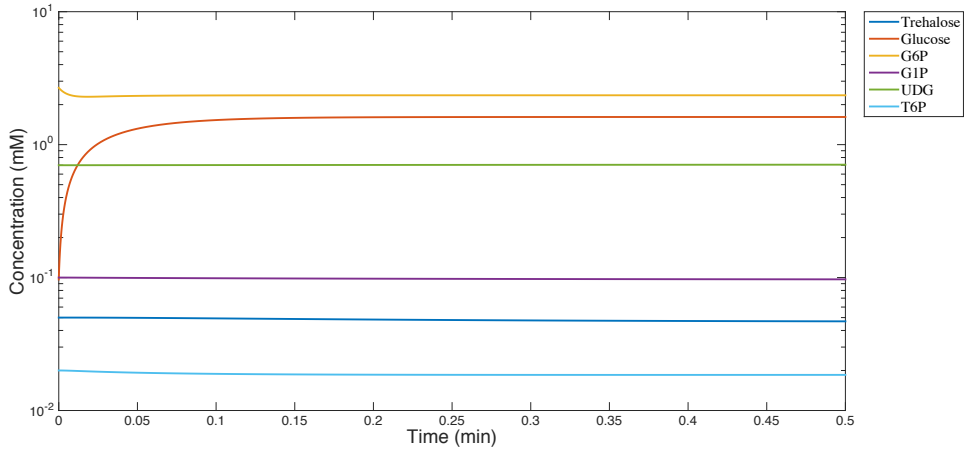


Figure 4.3.3. Modified Smallbone Model simulated using initial condition data of yeast (as used in the original Smallbone Model). Steady state values that are reached in this simulation are taken as the initial concentrations of Smallbone Modified Model in subsequent analyses.

The response of the modified model to heat shock (Figure 4.3.4), osmotic shock (Figure 4.3.5), TPS1 mutation (Figure 4.3.6), and trehalase mutation (Figure 4.3.7), was compared to the response of the original yeast model. Similar trends were observed (i.e. increase or decrease in the metabolite concentration, accumulation of trehalose in trehalase mutant), apart from the

concentration of UDG during heat shock, which is increased in the modified model and decreased in the yeast model. Table 4.3.12 lists the fold changes in concentration in the modified model.

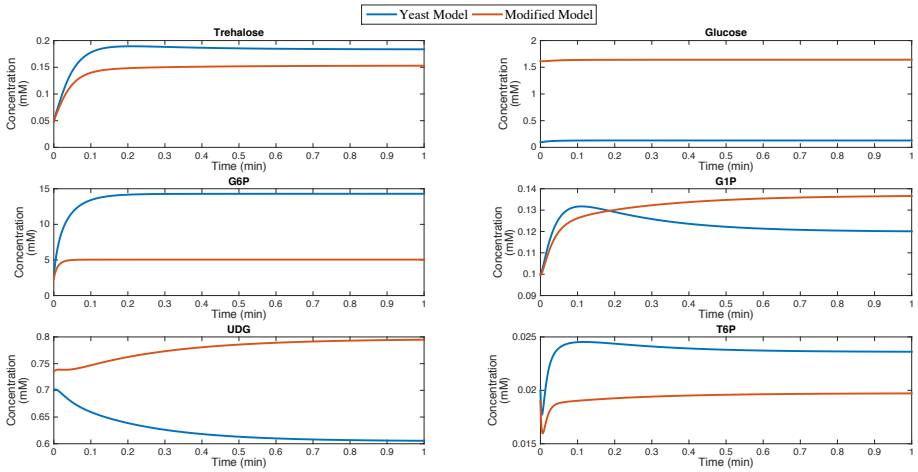


Figure 4.3.4. Metabolite concentrations in response to heat shock. Smallbone Model is marked in blue. Smallbone Modified Model is marked red. Changes in enzymatic activities induced by heat shock were taken from Smallbone et al. [203].

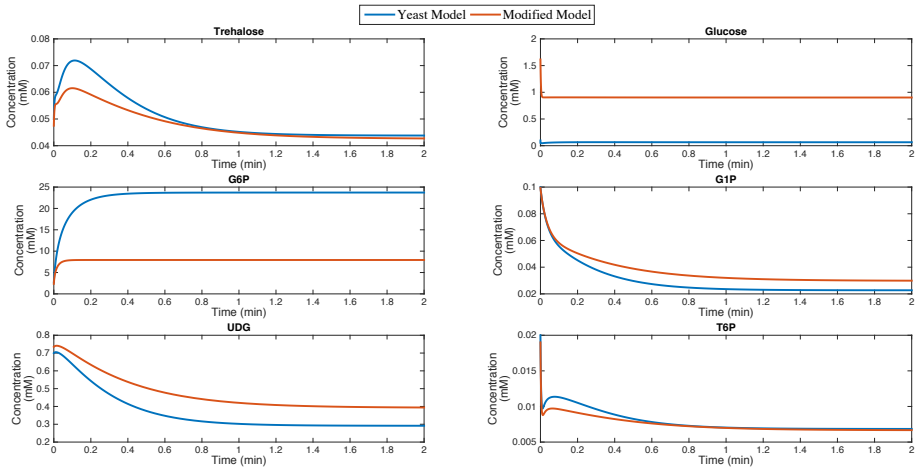


Figure 4.3.5. Metabolite concentrations in response to osmotic shock. Smallbone Model is marked in blue. Smallbone Modified Model is marked red. Changes in enzymatic activities induced by osmotic shock were taken from Smallbone et al. [203].

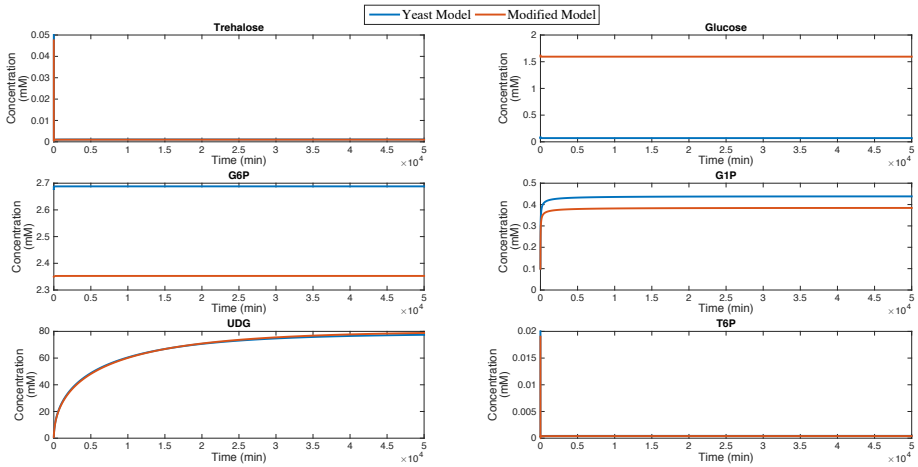


Figure 4.3.6. Metabolite concentrations in response to TPS1 mutantion. Smallbone Model is marked in blue. Smallbone Modified Model is marked red. TPS1 mutation was simulated by reducing TPS1 activity to 1% of the normal level.

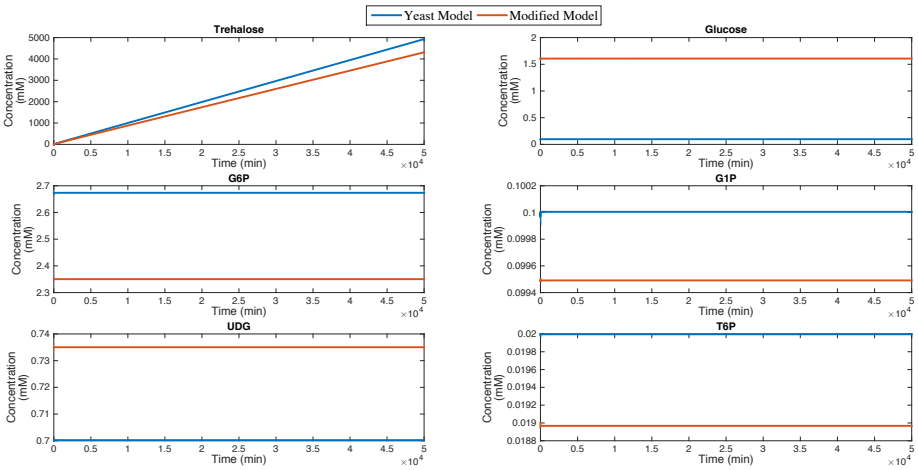


Figure 4.3.7. Metabolite concentrations in response to trehalase mutation. Smallbone Model is marked in blue. Smallbone Modified Model is marked red. Trehalase mutation was simulated by reducing trehalase activity to 1% of the normal level.

When the trehalase activity was reduced to 1% of its normal level (Figure 4.3.7), representing the trehalase mutant, the system cannot reach equilibrium. A functioning trehalose cycle is necessary to stabilise the system. When the link between the trehalose and glucose is impaired, trehalose accumulates. Accumulation of trehalose can be harmful for the cell. A simulation was carried out, where the trehalase activity was reduced to 10% of its normal level. This can represent a significant down-regulation of the trehelase gene. The system was stable with a significant increase in trehalose concentration (Figure 4.3.8).

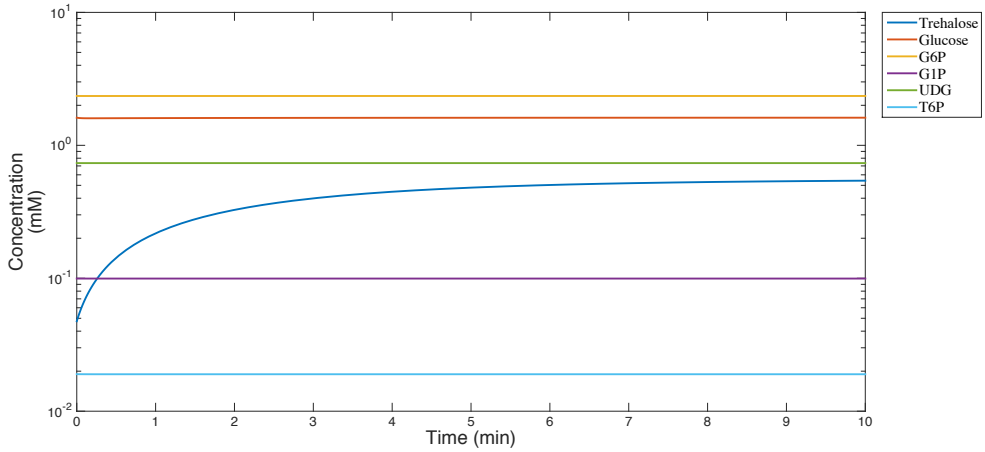


Figure 4.3.8. Smallbone Modified Model metabolite concentrations in response to down-regulation of trehalase (Trehalase activity is reduced to 10% of the normal level).

Table 4.3.12. Simulated fold change in the Smallbone Modified Model concentrations in responses to heat shock, osmotic shock, TPS1 mutation, and trehalase mutation. The values shown correspond to $\log_{10}([X]_{ss}/[x]_0)$, where $[X]_{ss}$ is the steady state metabolite concentration and $[X]_0$ is the initial metabolite concentration.

Metabolite	Heat shock	Osmotic shock	TPS1 mutant	Trehalase mutant (trehelase activity reduced to 1% of its normal level	Trehalase down-regulation (trehelase activity reduced to 10% of its normal level
Glucose	0.0073	-0.2527	-0.0053	-2.00E-03	-4.96E-06
Glucose 6-phosphate	3.32E-01	5.28E-01	3.53E-04	-8.06E-05	-4.33E-06
Glucose 1-phosphate	1.39E-01	-5.24E-01	0.5868	-3.90E-05	-3.38E-05
UDP glucose	0.0349	-0.2723	2.0366	1.76E-05	-1.09E-05
Trehalose 6-phosphate	0.0165	-0.4555	-1.6776	-7.41E-04	-7.08E-04
Trehalose	0.5083	-0.0474	-1.668	50.2556	1.0668

As reviewed in Section 4.2.4, genes such as trehalase and trehalose-phosphate synthase were found to be differentially regulated during infection [126,202]. The change in the encoding gene expressions (Table 4.2.1 and Table 4.2.2) were incorporated to the model and used to estimate the change in the activities of the enzymes to evaluate the metabolite concentrations in response to infection. Table 4.3.13 summarises the changes in metabolite concentrations in each condition.

Table 4.3.13. Simulated fold change in the Smallbone Modified Model concentrations in responses to infection. The values shown correspond to $\log_{10}([X]_{ss}/[x]_0)$, where $[X]_{ss}$ is the steady state metabolite concentration and $[X]_0$ is the initial metabolite concentration.

Metabolite	*Burkholderia pseudomallei* K96243 wild type Bp-infected murine lungs/ *Burkholderia pseudomallei* K96243 wild type	*B. pseudomallei* during intracellular growth in host macrophages relative to in vitro growth			
		1h	2h	4h	6h
Glucose	0.0403	-0.0053	-0.0047	-0.0051	-4.10E-03
Glucose 6-phosphate	-0.0052	51.1386	0.4359	1.2431	2.78E-01
Glucose 1-phosphate	0.1319	50.5063	0.9784	1.8034	0.7705
UDP glucose	0.0302	-0.2528	-0.2729	-0.5257	0.2123
Trehalose 6-phosphate	-0.2768	-1.5306	-0.905	-1.1818	-0.63
Trehalose	-0.3552	-1.521	-0.8964	-1.1726	-0.6224

The up-regulation of trehalase in infected murine lungs, results in slight down-regulation of the upstream metabolites (trehalose and trehalose 6-phosphate) and slight up-regulation of Glucose 1-phosphate (Figure 4.3.9).

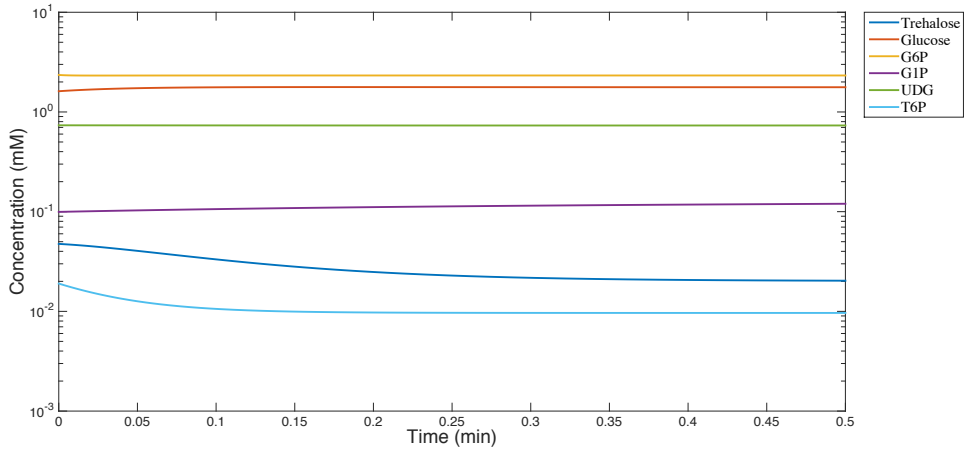


Figure 4.3.9. Change in Smallbone Modified Model concentrations in *B. pseudomallei* in infected murine lungs relative to in vitro growth.

During the first hour of infection in host macrophages, the system is not in equilibrium as glucose 6-phosphate and glucose 1-phosphate accumulate (Figure 4.3.10).

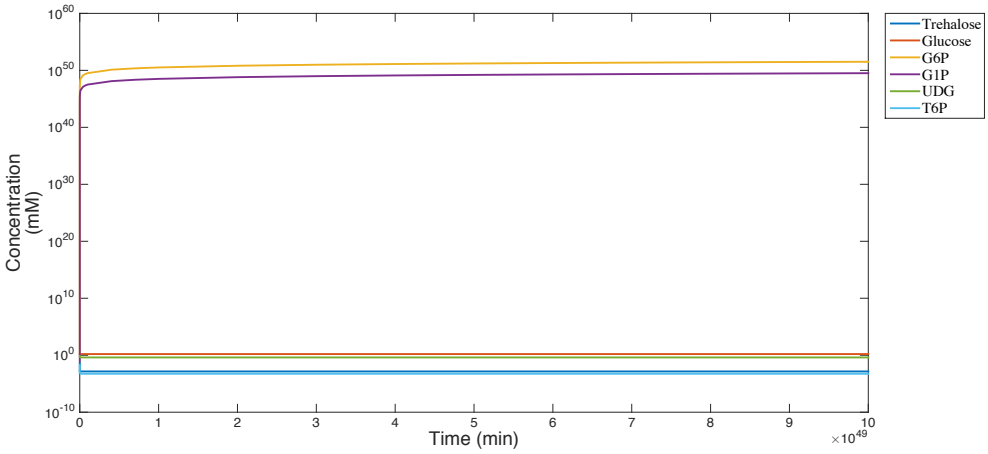
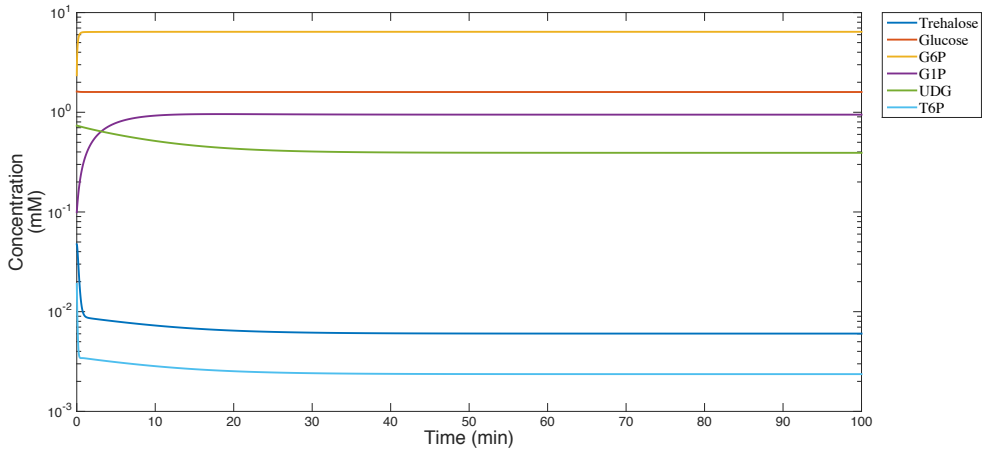


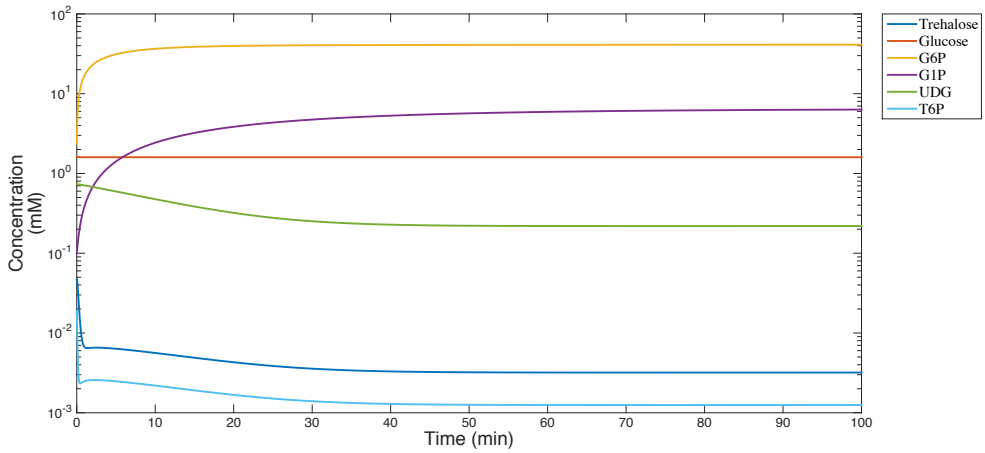
Figure 4.3.10. *B. pseudomallei* during intracellular growth in host macrophages relative to in vitro growth 1h

Substantial decrease in trehalose and trehalose 6-phosphate were observed in *B. pseudomallei* during 2h, 4h, and 6h in host macrophages (Figure 4.3.11).

A



B



C

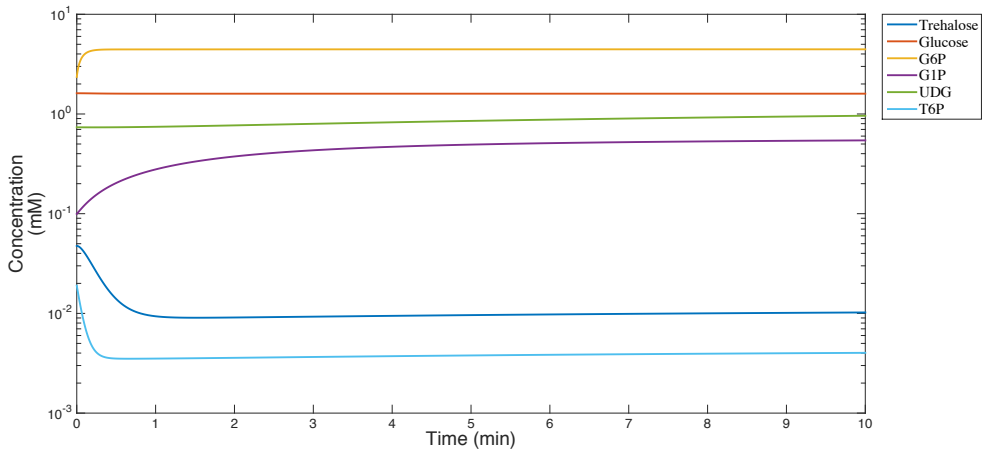


Figure 4.3.11. Change in Smallbone Modified Model concentrations in *B. pseudomallei* during intracellular growth in host macrophages relative to in vitro growth (A) 2h; (B) 4h; (C) 6h.

4.3.4 Sensitivity Analysis

Section 4.3.3 demonstrated the response of metabolite concentrations to mutation, stress, or infection. During mutation, the activity of the corresponding reaction was abolished. During infection, the activities of corresponding reactions were affected with the changes in gene expression. Stress survival and adaptation to the environmental changes can require regulation of various parts of the metabolic process. This regulation can be different in different organisms. Due to the complex organisation of metabolic networks, it is not straightforward to tell which metabolites will be affected when a parameter is changed. A sensitivity analysis was performed to determine how every single parameter in the model could impact the level of metabolites. In Figure 4.3.12, the sensitivities for the original and modified Smallbone models were plotted. The concentration of trehalose is sensitive to the parameters of the trehalase reaction. Thus a mutation in trehalase enzyme can significantly alter the trehalose levels in the cell. When the parameter of a reaction is changed, the levels of the substrates of that reaction are affected. Sensitivities are similar in the original and modified Smallbone Model with the exception of glucose; unsurprisingly glucose is more sensitive in the original model to changes in glucose import, hexokinase, T6P synthase and T6P synthase phosphatase because of the inhibitory effect of T6P on glucose consumption included in the original model. However, glucose 6-phosphate exhibited low sensitivity to small perturbations in T6P-related reaction parameters, though previous simulations showed an increase in glucose 6-phosphate concentration when these parameters were altered significantly (e.g. Figure 4.3.1A). The efficacy of hexokinase inhibition by T6P was challenged previously [226,235], postulated to be significant in a narrow metabolic range and only relevant in a glucose

pulse. Next, in Section 4.3.5 a kinetic model of glycolysis was studied to provide a better understanding of the role of the trehalose metabolism and its interactions with glycolysis.

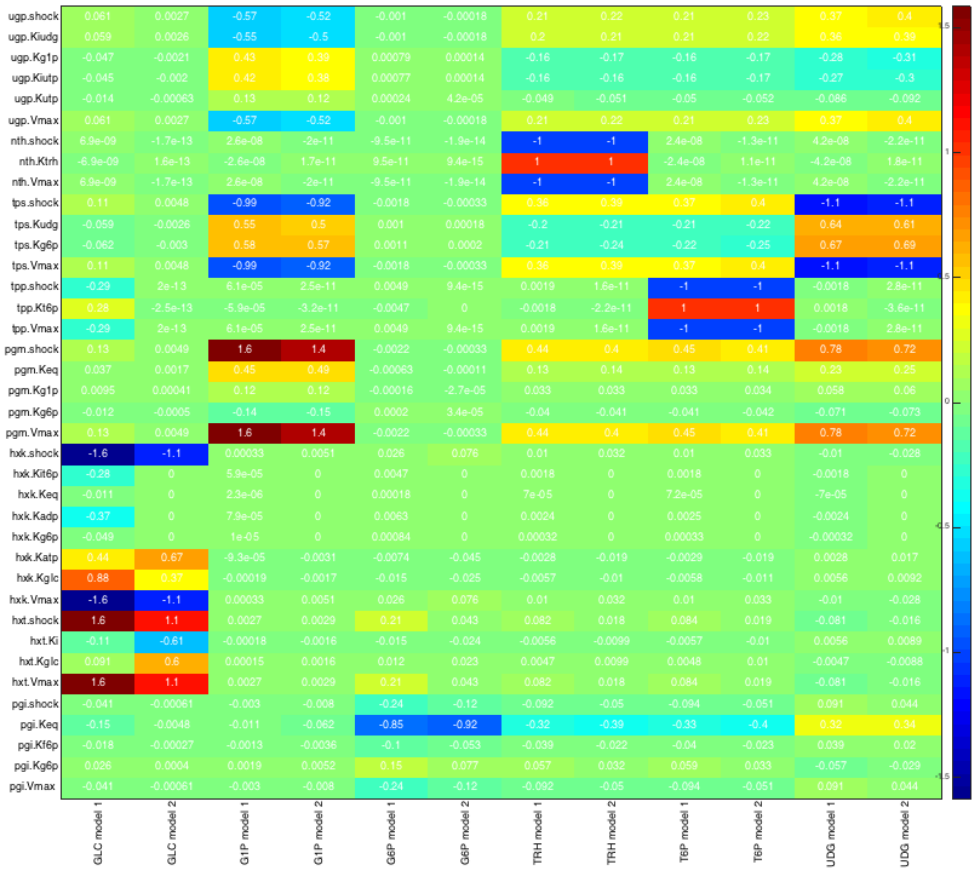


Figure 4.3.12. Smallbone and Modified Model sensitivity statistics of the steady states with respect to 1% parameter perturbations. The Smallbone Model is referred to as 'Model 1' and the Modified Smallbone Model is referred to as 'Model 2'. The rows correspond to the parameters, and the columns to metabolites. The colours (and the values on the cells) represent the sensitivity of the metabolite for the given parameter (calculated as explained in Section 4.2.5). Metabolites: GLC, glucose; G1P, glucose 1-phosphate; G6P, glucose 6-phosphate; TRH, trehalose; T6P, trehalose 6-phosphate; UDG, UDP-glucose. Parameters: ugp, UDP glucose phosphorylase; nth, trehalase; tps, T6P synthase; tpp, T6P phosphatase; hxk, hexokinase/glucokinase; hxt, glucose transport pgi, G6P isomerase; shock parameters represent the effect of change in the activity of the associated reaction.

4.3.5. Interaction of the trehalose cycle with glycolysis

Van Heerden *et al.* [204] demonstrated that the trehalose cycle in yeast helps maintain functional glycolysis and facilitates yeast in coping with stresses such as excess glucose availability. In yeast, on excess glucose, the flux through the upper glycolysis pathway (ATP-consuming) is greater than the flux through the lower glycolysis pathway (ATP-generating). This imbalance can impair growth. Figure 4.3.13 illustrates the upper- and lower glycolysis and the trehalose cycle forming a feedback loop to the upper glycolysis pathway by inhibiting the synthesis of glucose-6-phosphate (inhibition of hexokinase by T6P as discussed in previous sections). This feedback was found to prevent the accumulation of the intermediates of the upper-glycolysis pathway and maintain the normal functional state of glycolysis. Mutants with a defect in the trehalose cycle can end up in an imbalanced state, resulting in metabolic malfunctioning. An alternative mechanism through which the trehalose cycle can ensure the proper functioning of glycolysis is to recover phosphate when needed. The trehalose cycle, as illustrated in Figure 4.2.1, produces phosphate during synthesis of UDP-glucose and trehalose.

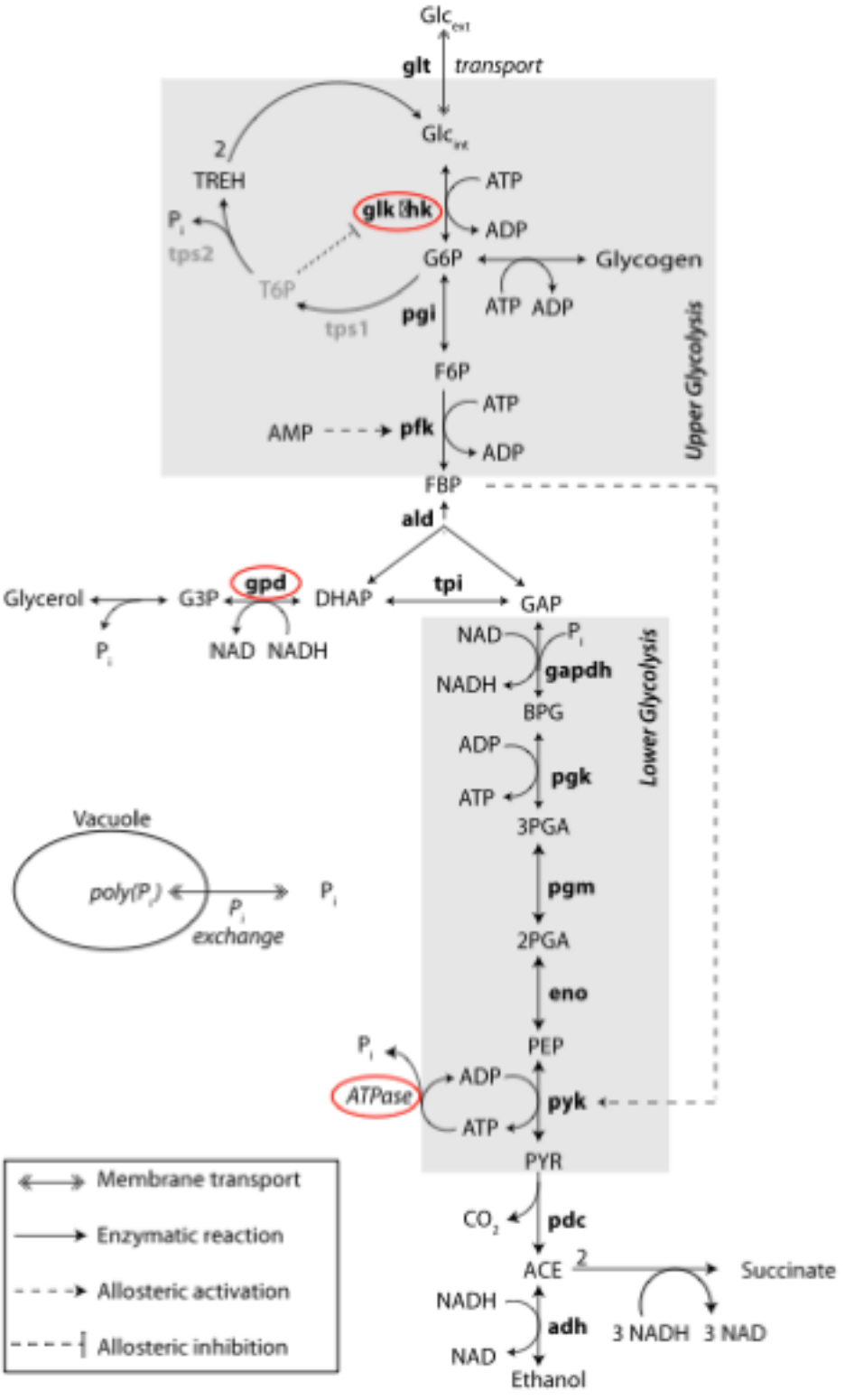


Figure 4.3.13. The trehalose cycle and glycolysis in yeast. Figure adapted from [204]

This section first reviews the wild type and *TPS1* mutant behaviour of the van Heerden Model. The results published by van Heerden *et al.* were reproduced and the analysis is expanded to include the trehalase mutant.

Van Heerden *et al.* adapted the yeast kinetic glycolysis model of Teusink *et al*. [32] and introduced modifications that represents the interactions of the trehalose cycle with glycolysis. Of note, the trehalose cycle reactions are not defined explicitly. Instead, as the magnitude of change in G6P was found to be similar to that of T6P and trehalose [236], G6P levels were used to model the T6P inhibition and the flux from trehalose toward glycolysis, via.

$$\frac{d\,[G6P]}{dt} = f1 - f2 - f3 + f4$$

In the above f1 is the flux through the hexokinase reaction and the rate equation of hexokinase (Equation) includes a term for the T6P inhibition of hexokinase (highlighted in yellow; $K_{i,G6P}$, T6P inhibition constant of hk)

$$v_{hk} = \frac{v_{max,hk}\left(\frac{[ATP]}{K_{ATP}}\frac{[GLCi]}{K_{GLCi}} - \frac{[ADP]}{K_{eq,hk}K_{ATP}}\frac{[G6P]}{K_{GLCi}}\right)}{\left(1 + \frac{[ADP]}{K_{ADP}} + \frac{[ATP]}{K_{ATP}}\right)\left(1 + \frac{[G6P]}{K_{i,G6P}} + \frac{[G6P]}{K_{G6P}}\frac{[GLCi]}{K_{GLCi}}\right)}$$

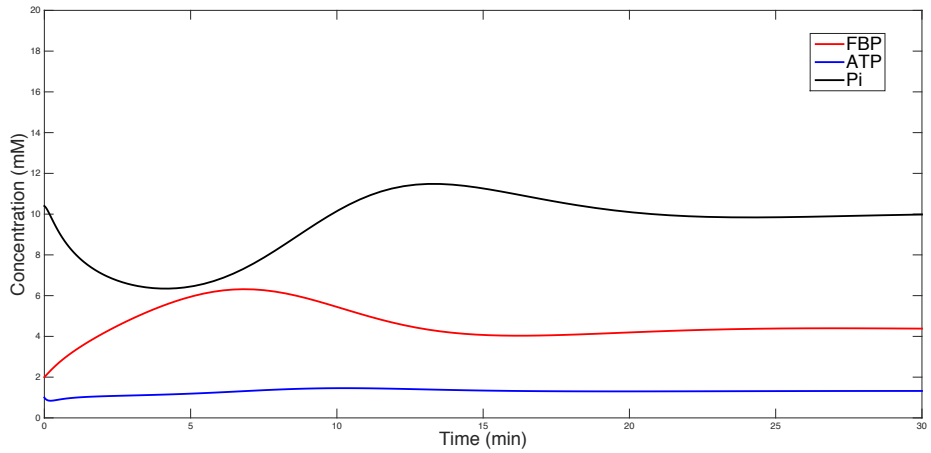
f2 is the flux towards F6P; f3 is the flux towards glycogen;

and f4 is the flux feeding to glycolysis from trehalose represented as $k_{trehalose} * G6P$.

Furthermore, as the trehalose cycle was postulated to function as a phosphate recovery mechanism, phosphate was included as a free variable in the model. Whether the balanced state is reached depends on the initial concentration of the metabolites. When the initial concentration of phosphate is 10.4 mM, the system reaches a steady state (Figure 4.3.14A). The difference in

rate between the upper and lower glycolysis pathways is zero (Figure 4.3.14B). Thus, there is no accumulation of metabolites and the system is in a balanced state.

A



B

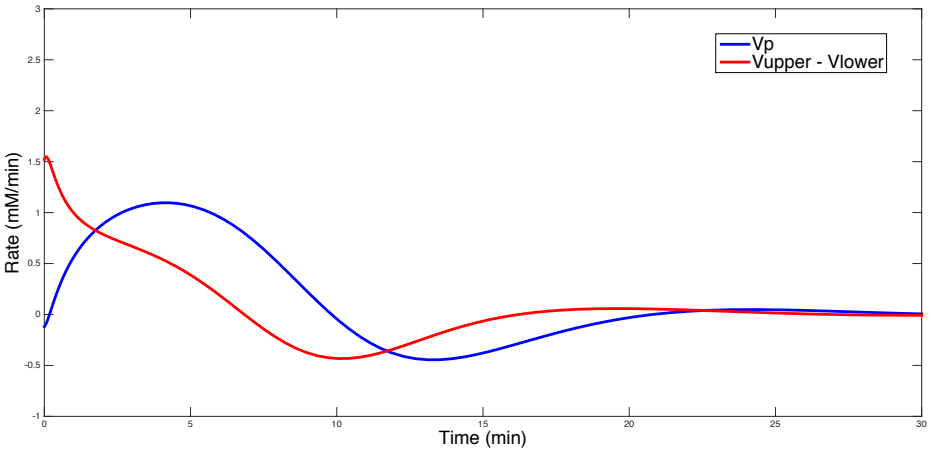
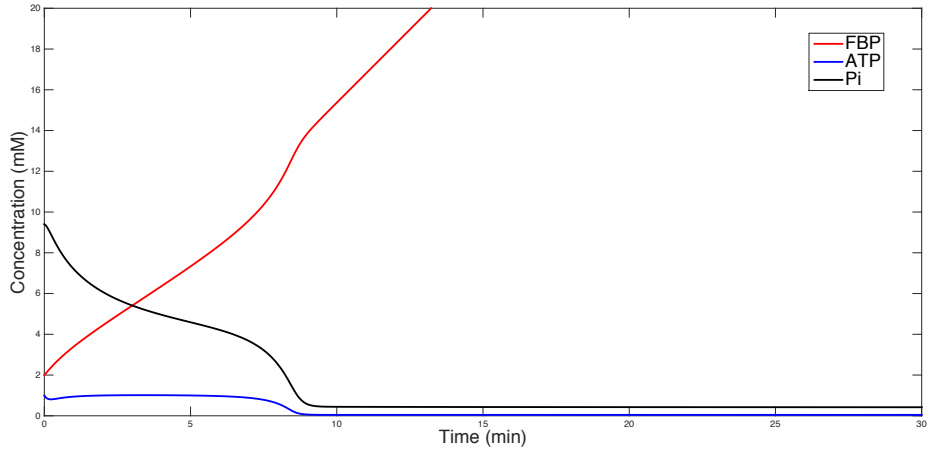


Figure 4.3.14. Van Heerden Model simulated with initial Phosphate (Pi) level of 10.4 mM. FBP, fructose 1,6-biphosphate; ATP, adenosine 5'-triphosphate; Pi, cytosolic phosphate; v_p, the vacuolar import rate of Pi; v_{upper}, the rate of upper glycolysis; v_{lower}, the rate of lower glycolysis. (A) Metabolite concentrations reaching steady state. (B) The vacuolar import rate of Pi (vp) reaches zero at steady state. The rate between upper and lower glycolysis ($v_{upper} - v_{lower}$) reaches zero. These simulations reproduce Fig. 4 (C and E) in van Heerden *et al.* [209].

When the initial concentration of phosphate is 9.4 mM, the glycolytic intermediate fructose 1,6-bisphosphate (FBP) is accumulates (Figure 4.3.15 A). Upper glycolysis is faster than lower glycolysis and vacuolar transport of phosphate is not at steady state (Figure 4.3.15B). This imbalance impairs ATP generation and growth. This shows that phosphate dynamics are a key feature of the network.

A



B

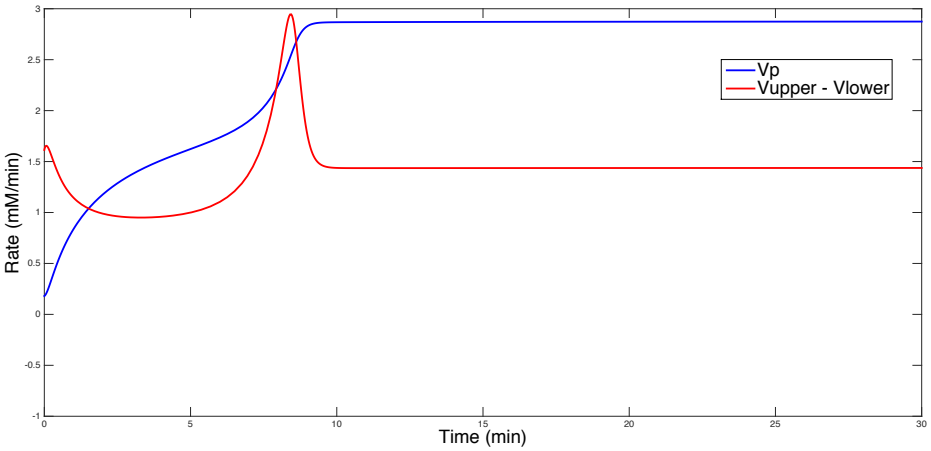


Figure 4.3.15. Van Heerden Model simulated with an initial phosphate (Pi) level of 9.4 mM. FBP, fructose 1,6-biphosphate; ATP, adenosine 5'-triphosphate; Pi, cytosolic phosphate; v_p, the vacuolar import rate of Pi; v_{upper},

the rate of upper glycolysis; v_{lower}, the rate of lower glycolysis. (A) Metabolite concentrations resulting in imbalance (FBP accumulation and low ATP and Pi levels). (B) The vacuolar import rate of Pi (vp) and the rate between upper and lower glycolysis ($v_{upper} - v_{lower}$) is positive. These reproduce Fig. 4 (D and F) in van Heerden *et al.* [209].

The imbalanced and the balanced state also depend on the feedback from T6P (trehalose 6-phosphate) to hexokinase (hxk). In the wild type there is no accumulation of metabolites (Figure 4.3.16).

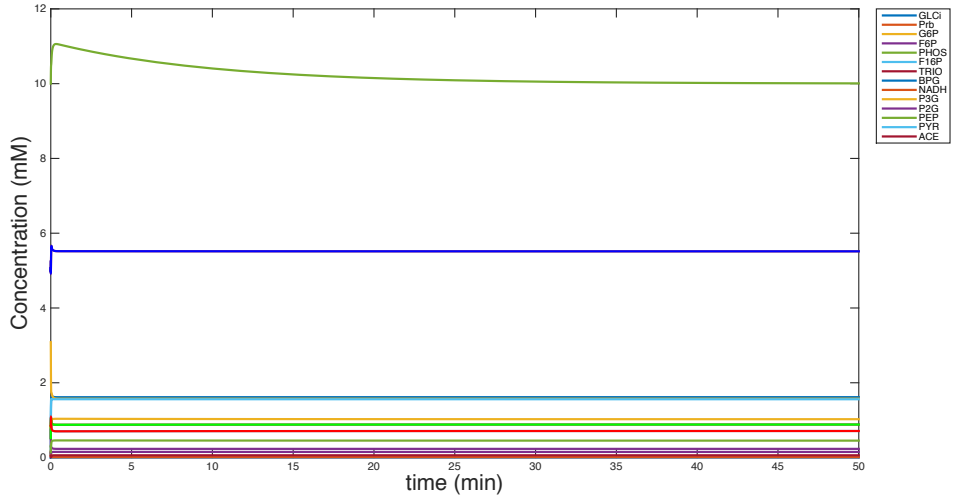


Figure 4.3.16. Simulated wild type metabolite concentrations in the van Heerden Model.

When the feedback on hk is eliminated and the rate constant of trehalose branch is set to zero, representing the *TPS1* mutant as defined by van Heerden *et al.*, FBP accumulates while ATP and phosphate are depleted (Figure 4.3.17).

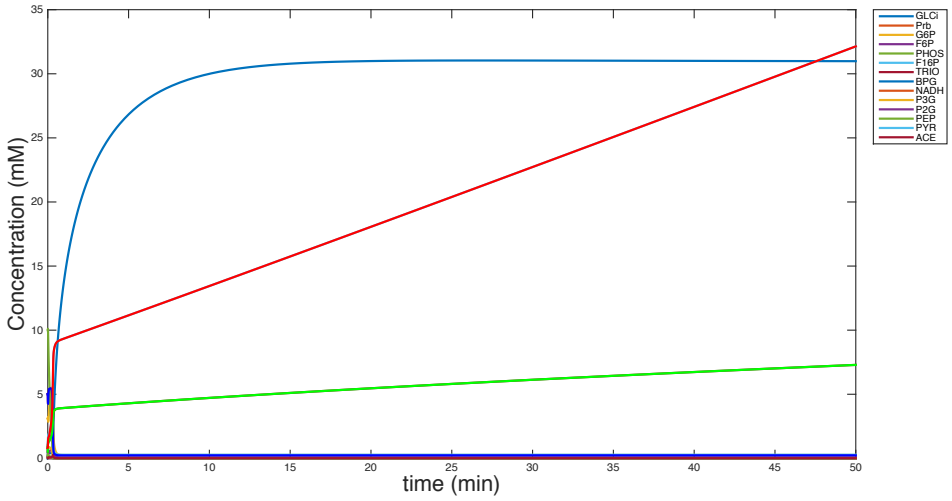


Figure 4.3.17. Metabolite concentrations in van Heerden Model in TPS1 mutation. TPS1 mutation (as defined by van Heerden *et al.*) is simulated by eliminating the feedback on hexokinase and setting the trehalose branch flux to zero.

In simulations, van Heerden *et al*. characterised the *TPS1* mutant with the complete malfunction of the trehalose cycle, as all interactions between the trehalose cycle and glycolysis were eliminated to simulate the mutant phenotype (combined elimination of trehalose cycling and T6P-mediated feedback on hexokinase).

In this section, the effects of eliminating the feedback on hexokinase and eliminating the trehalose branch were considered separately. To mimic a trehalase mutation, the rate constant of the trehalose branch was set to zero, while the feedback on hexokinase was kept intact. In this case, proper glycolytic functioning was established, as shown in Figure 4.3.18 - no accumulation of metabolites was observed.

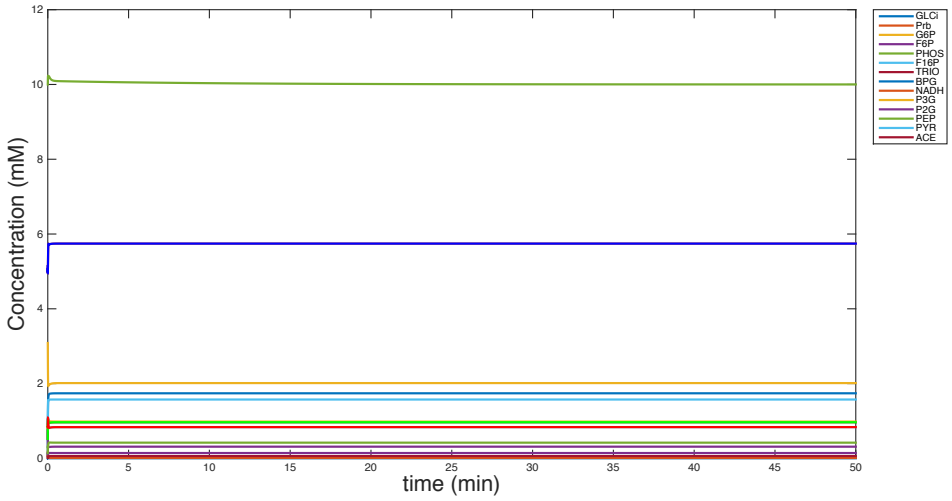


Figure 4.3.18. Trehalase mutation. Simulated metabolite concentrations in the van Heerden Model under trehalase mutation (the trehalose branch flux was set to zero

Next, a simulation was conducted in which the T6P-mediated feedback on hexokinase was eliminated and trehalose cycling was allowed. Trehalose cycling alone resulted in a functional glycolysis, where all metabolites were balanced and the phosphate level was recovered (Figure 4.3.19).

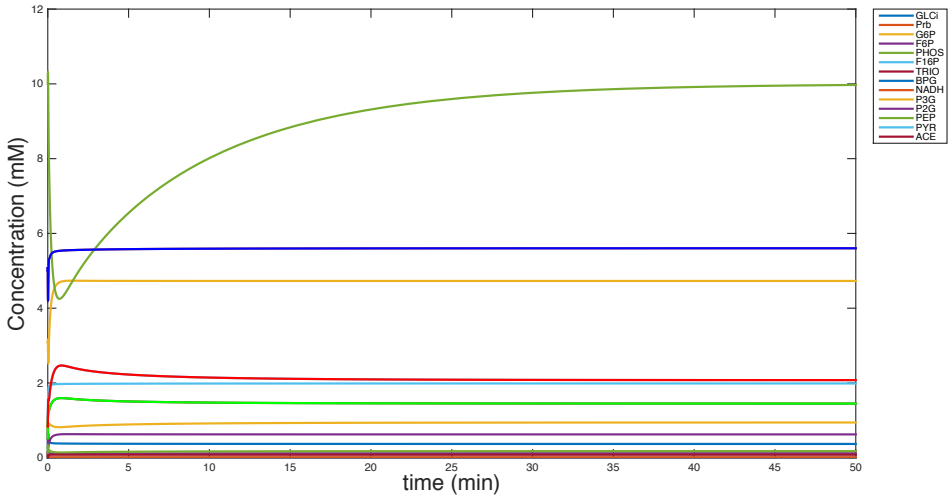


Figure 4.3.19. Simulated model concentrations in the van Heerden Model, when the feedback on hexokinase is eliminated.)

The simulations show that balance in glycolysis is affected only when the trehalose cycle is shut down completely. As discussed in Section 4.3.1, in bacteria, the reaction from glucose to glucose 6-phosphate is mediated by glucokinase, for which no T6P mediated inhibitory effects were observed. As this aspect of trehalose metabolism is likely to be absent in bacteria, eliminating the trehalose cycling alone (trehalase mutant) can result in malfunctioning of bacterial glycolysis.

4.4 Genome-scale metabolic network of *B. pseudomallei K96243*

In *B. pseudomallei*, reactions around trehalose are not limited to the ones in the trehalose cycle. Deciphering interactions, which *B. pseudomallei* trehalose may possess, is crucial for a complete understanding of the trehalase mutation. As the genome of *B. pseudomallei* is completely sequenced, metabolic networks of *B. pseudomallei* can be analysed at the genome scale to identify molecular links to trehalose.

B. pseudomallei K96243 is the first sequenced strain of *B. pseudomallei* [121]. The reference genome is 7,247,547 bp in size consisting of two chromosomes of 4,074,542 bp and 3,173,005 bp. Using the genomic data available, the metabolic network of *B. pseudomallei* K96243 was constructed using Model Seed [223]. The model contains a biomass reaction, which is used as an estimate of growth. A summary of the model is presented in Table 4.4.1.

Table 4.4.1. Summary of the *B. pseudomallei K96243* genome-scale metabolic model.

Organism	*Burkholderia pseudomallei K96243*
Model ID	Seed272560.3
Class	Gram negative
Genome size	7,247 KB
Model genes / Genes in genome	1,182 / 5,819
Reactions with genes	1,314
Exchange reactions	125
Compounds	1,143
Number of compounds in biomass reaction	81

To identify the reactions around trehalose in *B. pseudomallei*, the metabolites that are directly connected to trehalose were identified. This

subnetwork was extended by adding the connecting reactions and main reactions in glycolysis; the network structure is illustrated in Figure 4.4.1.

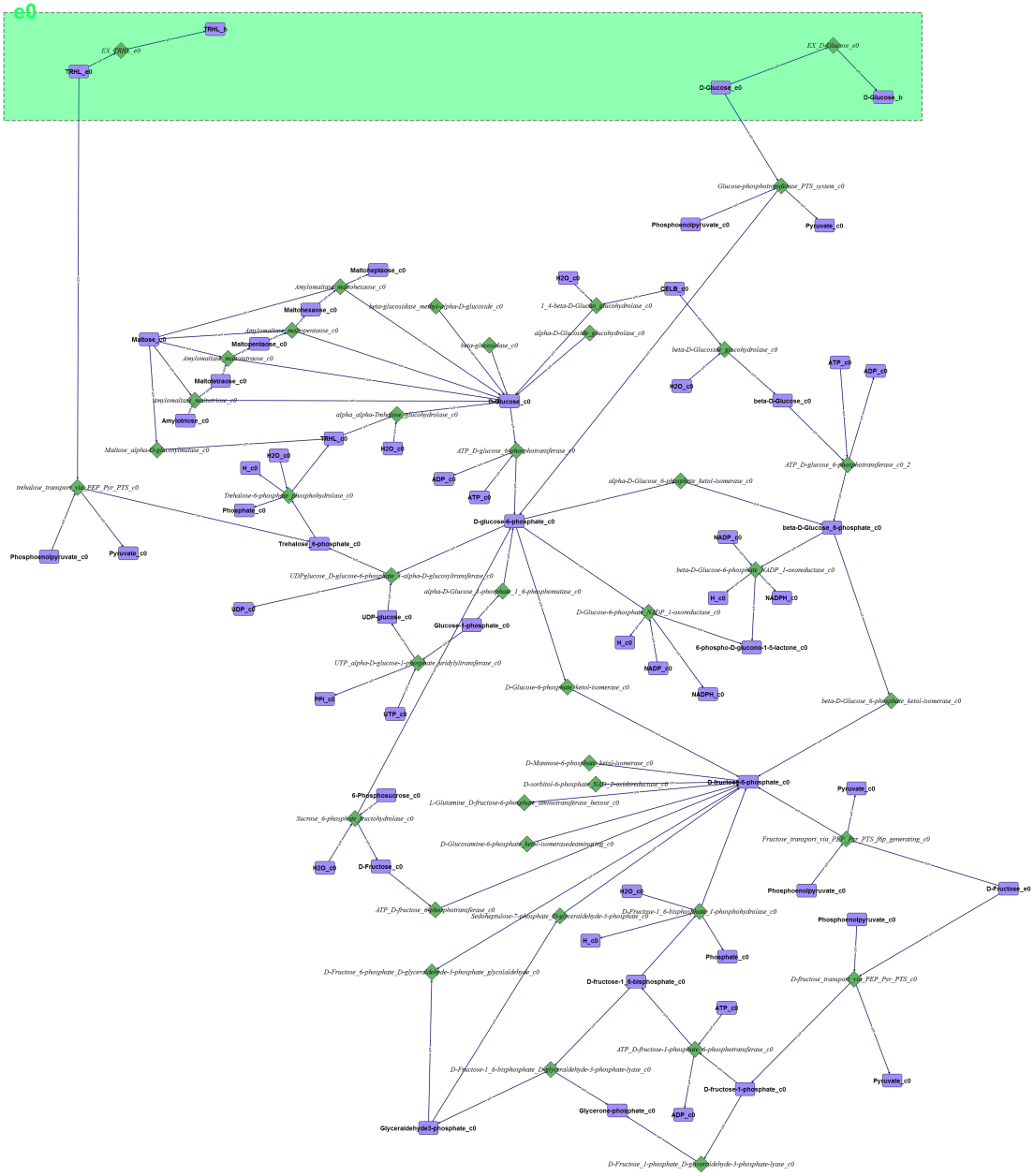


Figure 4.4.1. Subnetwork of the *B. pseudomallei* genome-scale metabolic model. Green area: extracellular compartment. Purple rectangles: metabolites. Diamonds: reactions.

Nine different simulations were run in which the maximum uptake rates of glucose and trehalose were changed on the defined minimal media composition required for growth. Negative values for the given carbon source means an intake to the cell.

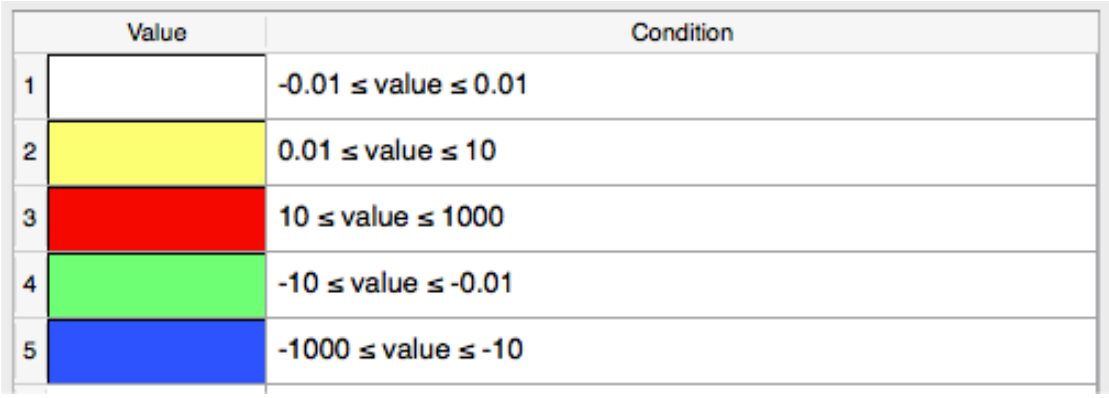
Biomass increased with an increasing concentration of glucose. The system was also able to grow on trehalose as there exists a trehalose uptake mechanism in the model. Although a few mechanisms have been suggested (e.g using periplasmic trehalose to convert trehalose into glucose followed by phosphotransferase-mediate glucose uptake [98]), the exact mechanisms of trehalose uptake are not fully understood.

Table 4.4.2. Growth simulations of the genome-scale metabolic model in different media compositions.

Simulation	D-Glucose	Trehalose	Biomass
sim1	-10	0	1.307
sim2	-50	0	6.5349
sim3	-100	0	13.0698
sim4	-1000	0	69.6929
sim5	0	0	4.69E-13
sim6	-10	-10	3.9209
sim7	-10	-100	27.4465
sim8	0	-10	2.614
sim9	0	-100	26.1395

For each simulation, the fluxes were mapped in the subnetwork using the color-coding explained in Figure 4.4.2. Positive and negative values for fluxes

determine the direction of flux in a reversible reaction (denoted by a double arrow). As an example, in Figure 4.4.3 the reaction “EX D_Glucose”, which is the uptake reaction of glucose, is a reversible reaction. It is coloured green in this simulation, hence the flux is in the opposite direction the (it is towards “D-Glucose”.



	Value	Condition
1		-0.01 ≤ value ≤ 0.01
2		0.01 ≤ value ≤ 10
3		10 ≤ value ≤ 1000
4		-10 ≤ value ≤ -0.01
5		-1000 ≤ value ≤ -10

Figure 4.4.2. Color-coding scheme used to represent fluxes in the flux maps shown in Figures 4.4.3-4.4.7. When the value is positive, the flux goes in the same direction as the arrow on the pathway figures (yellow – low flux rate; red – high flux rate). When the value is negative, the flux goes in the opposite direction to the arrow (green – low flux rate; blue – high flux rate). White, no flux.

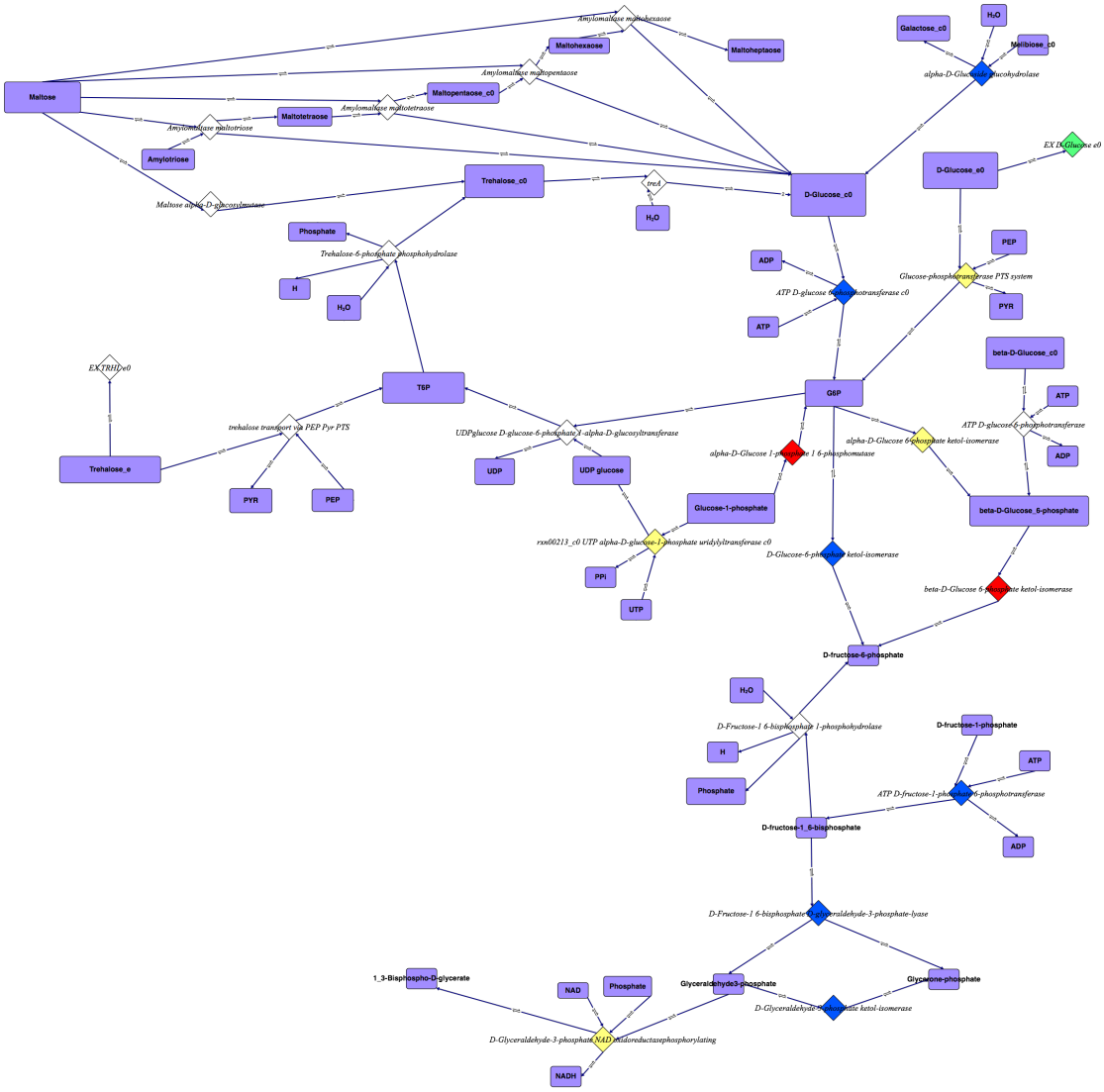


Figure 4.4.3. Sim 1. Subnetwork of the *B. pseudomallei* genome-scale metabolic model. Distribution of fluxes during growth in glucose: -10. The fluxes are represented using the color-coding scheme in Figure 4.4.2.

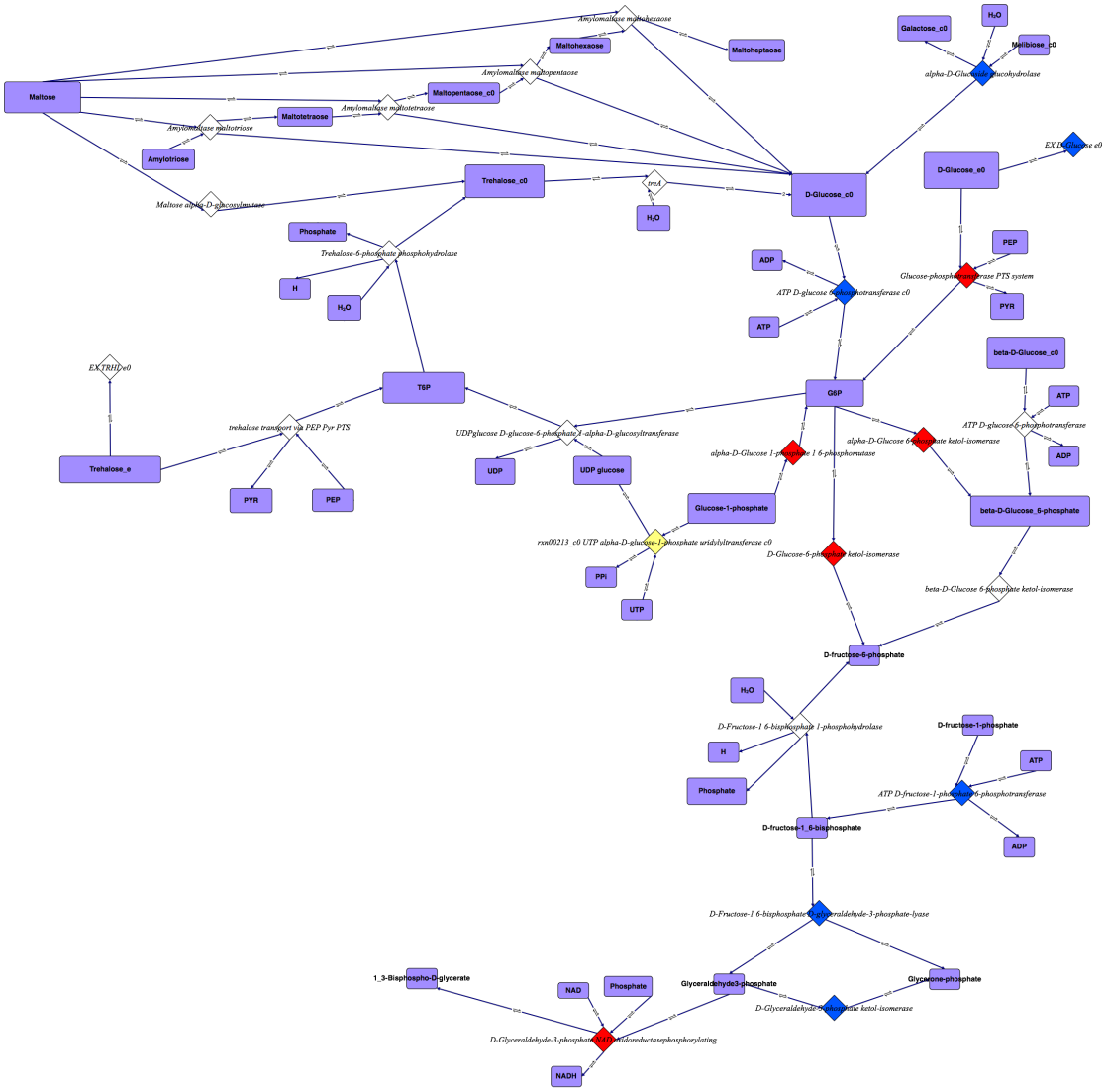


Figure 4.4.4. Sim 4. Subnetwork of the *B. pseudomallei* genome-scale metabolic model. Distribution of fluxes during growth in Glucose: -1000. The fluxes are represented using the color-coding scheme in Figure 4.4.2.

The flux maps (Figure 4.4.3 and Figure 4.4.4) show that as more glucose is available, more flux is going through certain reactions. However no flux is going through the trehalose cycle. The system optimizes biomass production and hence prefers direct routes leading to biomass precursors. For this, the trehalose cycle is not an efficient path for the system.

The flux maps in Figure 4.4.5 and Figure 4.4.6 show that the trehalase reaction doesn't carry flux even when there is trehalose available as the carbon source. Different objective functions can provide different flux distributions [237].

To demonstrate that the trehalose cycle can carry flux, a test objective function was generated, which optimises the production of trehalose. Under this condition, the trehalase reaction was indeed able to carry flux (Figure 4.4.7). This attests that, with formulating relevant objective functions, the model can be used to study fluxes around trehalose. Under stress conditions such as a heat shock, the cell might need to produce more trehalose.

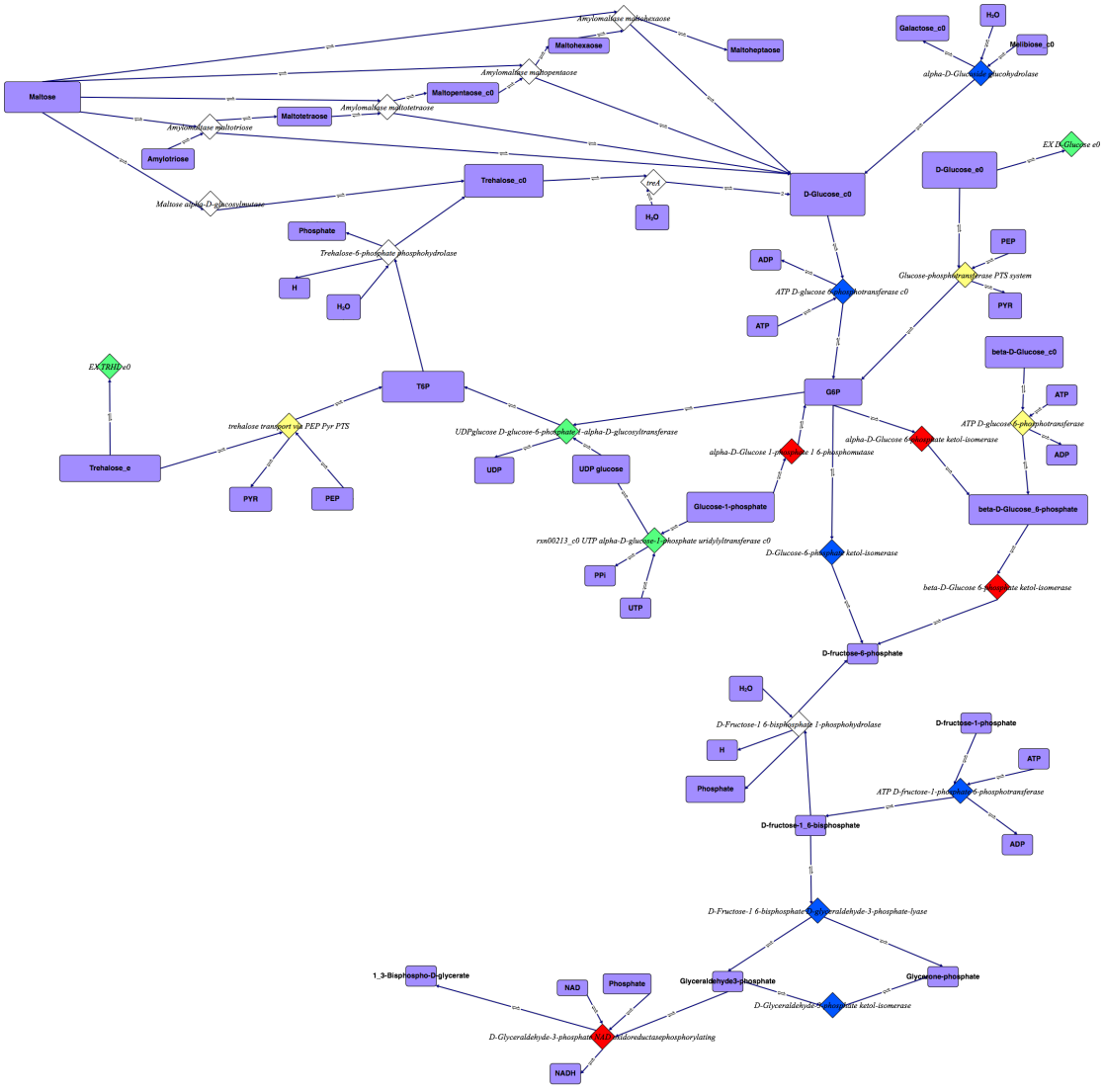


Figure 4.4.5. Sim 6. Subnetwork of the *B. pseudomallei* genome-scale metabolic model. Distribution of fluxes during growth in glucose: -10 and Trehalose: -10. The fluxes are represented using the color-coding scheme in Figure 4.4.2.

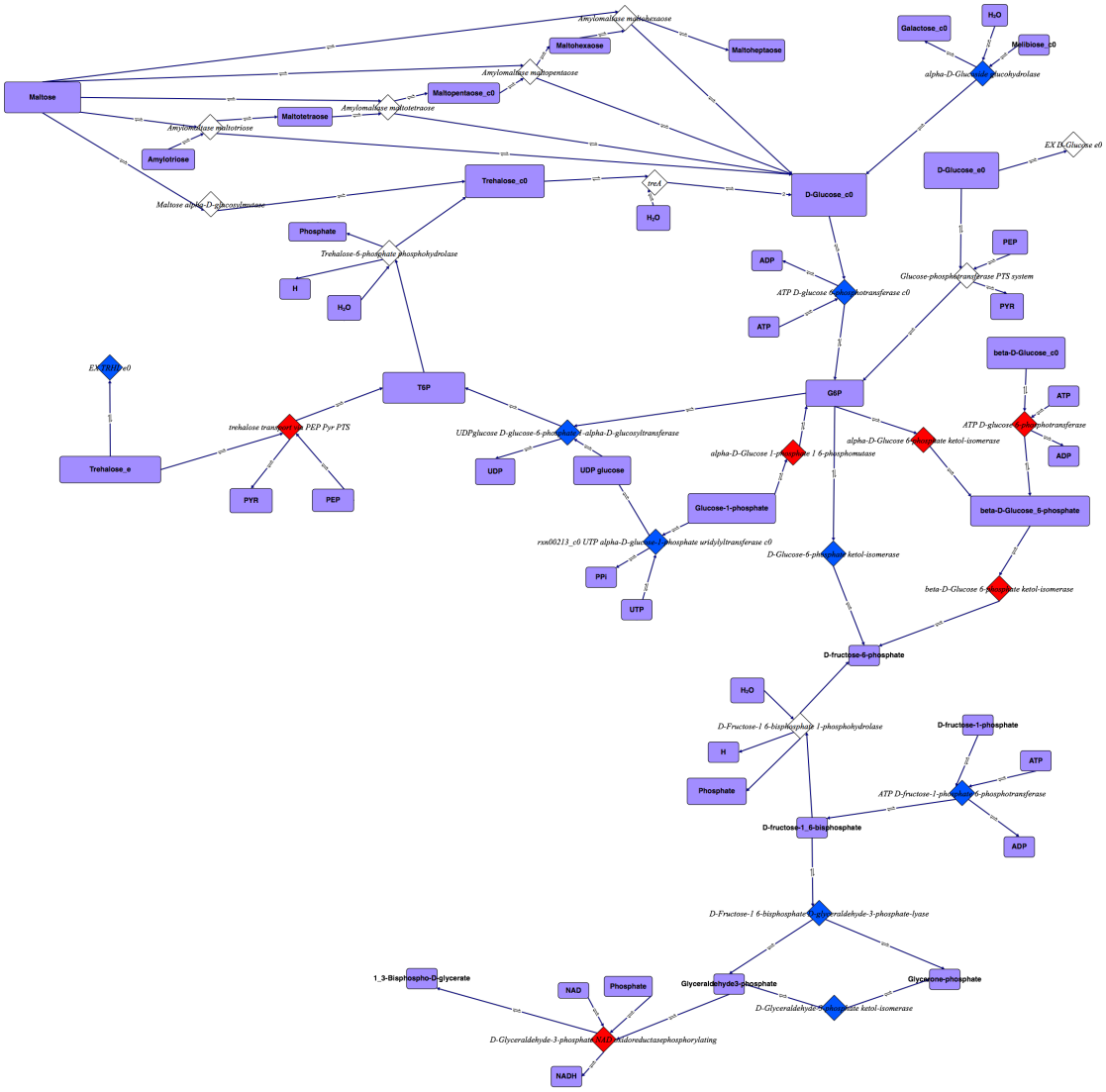


Figure 4.4.6. Sim 9. Subnetwork of the *B. pseudomallei* genome-scale metabolic model. Distribution of fluxes during growth in trehalose: -100. The fluxes are represented using the color-coding scheme in Figure 4.4.2.

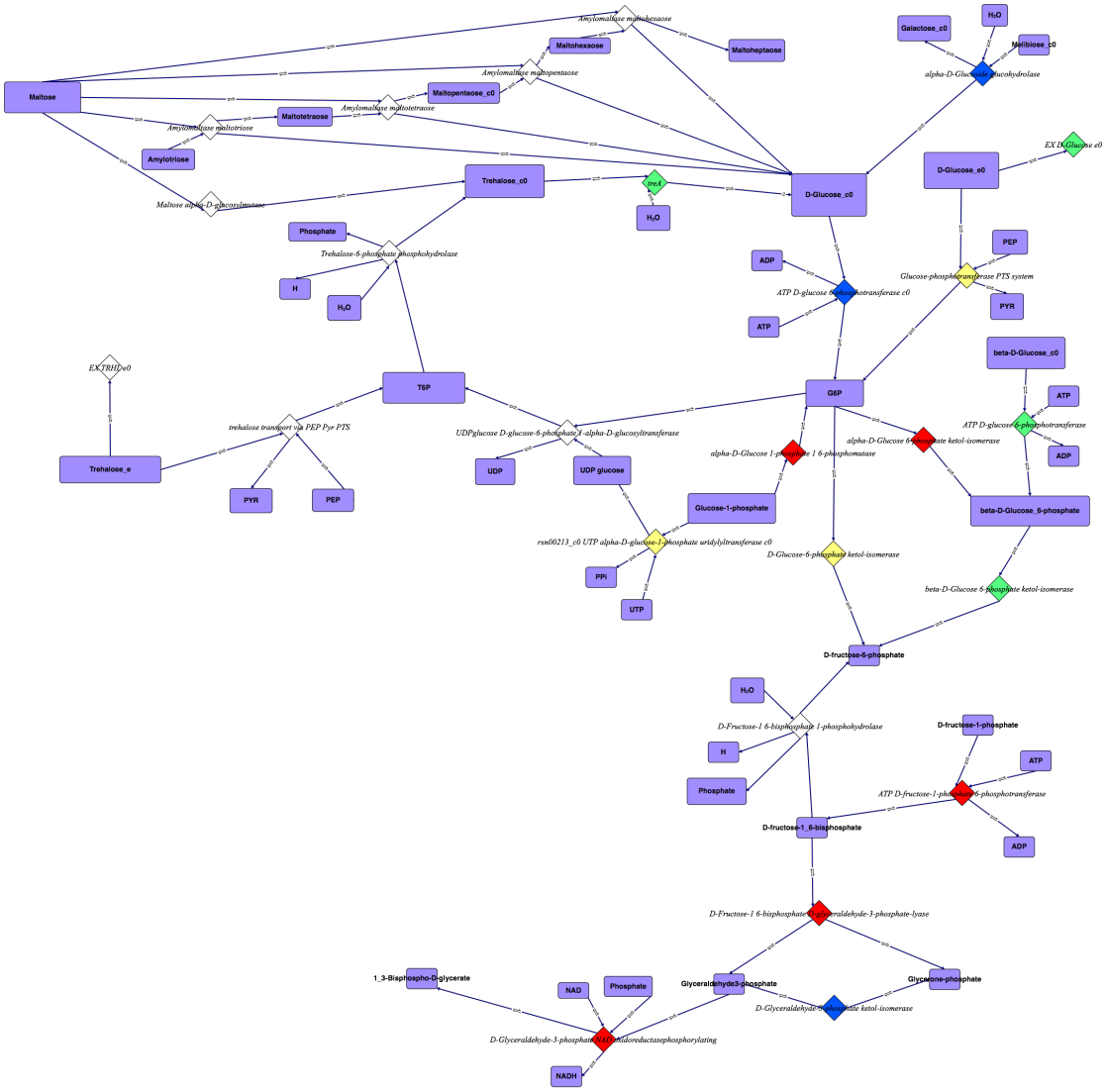


Figure 4.4.7. Subnetwork of the *B. pseudomallei* genome-scale metabolic model. Distribution of fluxes Flux distribution when trehalose is added to the objective function. The fluxes are represented using the color-coding scheme in Figure 4.4.2.

4.5 Functional analysis of the genome-scale metabolic network of *Burkholderia pseudomallei* K96243 and gene essentiality

The functions of the genes present in the model were analysed using the annotations available in *Burkholderia* Genome Database [212]. The genes were assigned a specific functional description using the Clusters of Orthologous Groups (COG) annotation. These functional descriptions are associated with more general category letters. The general categories were used to give an overview of the functional distribution of the genes (Fig 4.5.1). The majority of the genes were associated with metabolic functions such as amino acid (22.3%), coenzyme (10.2%), carbohydrate (7.7%) transport and metabolism.

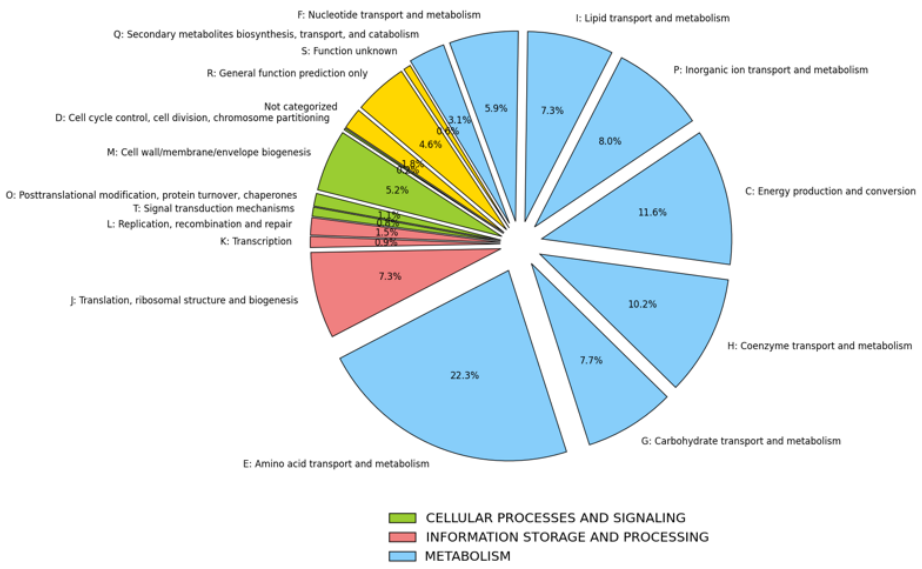


Figure 4.5.1. Functional distribution of genes (n = 1182) present in the genome-scale metabolic model of *Burkholderia pseudomallei* K96243, SEED272560_3. The genes were assigned a specific functional description using the Clusters of Orthologous Groups (COG) database. These functional descriptions are associated with general category letters.

To identify the essential bacterial genes required for bacterial growth, over 10^6 transposon insertion mutants were screened by Moule et al. using TraDIS [238]. TraDIS experiments were performed in Luria-Bertani (LB) broth or

agar at 37°C. Genes with no or few insertion sites were predicted to be essential. This resulted in 505 *Burkholderia pseudomallei* K96243 genes predicted to be essential for *in vitro* growth and survival. Although Moule *et al.* provided a functional analysis of the essential genes, for consistency with the previous functional analysis in Figure 4.5.1, the COG category distribution of these experimentally predicted essential genes were determined. The results were consistent with the Gene Ontology analysis presented in the study. The majority of the genes were associated with metabolic functions such as energy production, amino acid-, carbohydrate-, nucleic acid metabolisms (Figure 4.5.2).

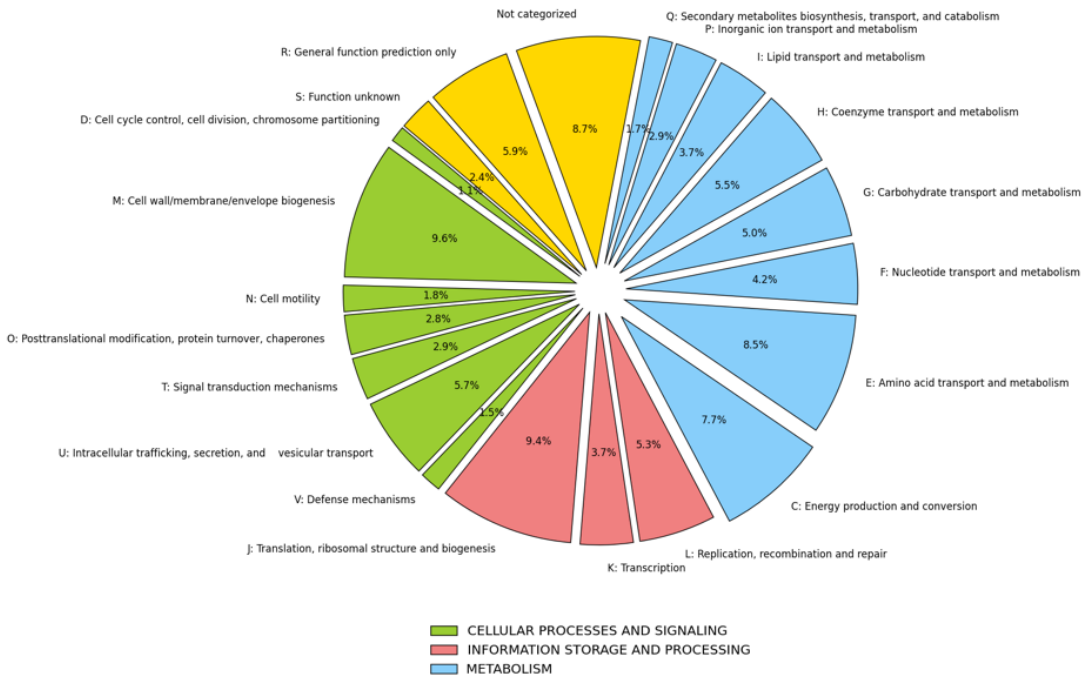


Figure 4.5.2. Functional distribution of essential genes predicted using TRADIS. The genes were assigned a specific functional description using the Clusters of Orthologous Groups (COG) database. These functional descriptions are associated with general category letters. The majority of genes were associated with metabolic functions. Other essential functions are translation, transcription, and transport.

Out of these 505 genes, 211 genes were present in the genome-scale model of *Burkholderia pseudomallei* K96243 SEED272560_3. Functional analysis was performed on the remaining 294 genes, to see whether there are any metabolic genes (Figure 4.5.3). 75 metabolic genes were found. The specific functionality of these genes are to be assessed and the genes can be potentially added to the metabolic model. The rest of the genes can be virulence factors providing an advantage to the pathogen in diverse environments such as in a host. However, these are less likely to be included in the metabolic models constructed to simulate *in vitro* growth.

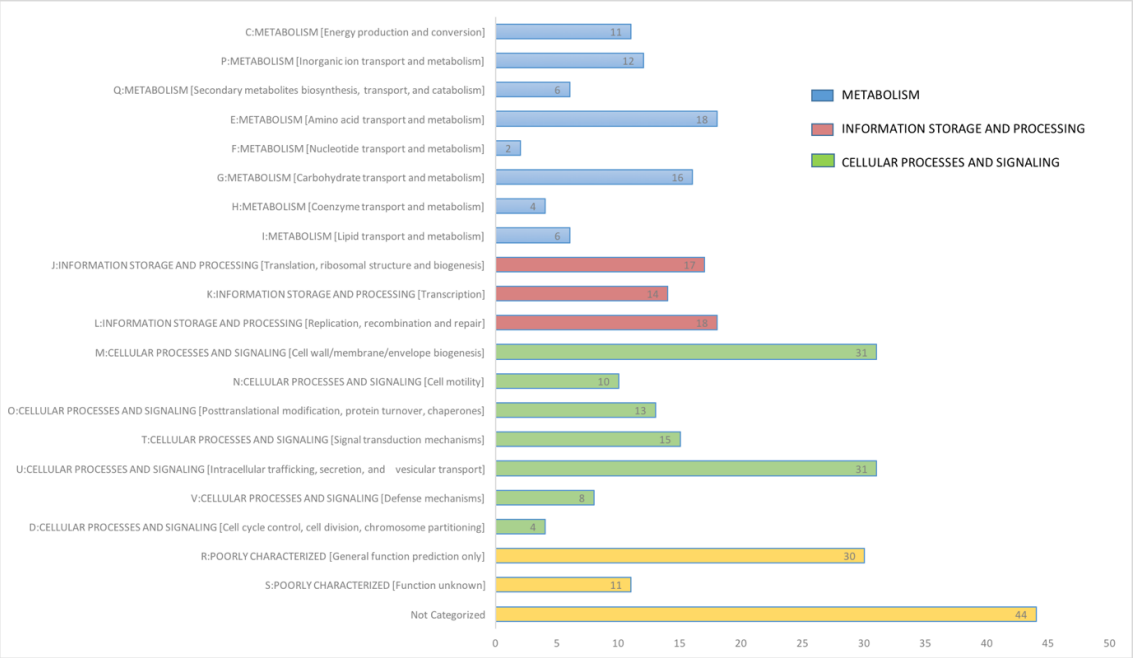


Figure 4.5.3. Functional distribution of essential genes predicted using TRADIS, but not present in the genome-scale metabolic model. The number of genes in each category is shown. There are 294 genes in total.

A single gene deletion analysis was performed on to the genome-scale model of *B. pseudomallei* K96243 SEED272560_3. Deletion of a gene forces fluxes through associated reactions to be zero. A gene is marked as essential

when maximum growth rate (biomass production) is zero upon the gene's deletion. Gene essentiality is analysed separately in different media compositions.

13 different media were chosen from the MODEL SEED. Media compositions are listed in Table 2. There was no growth in media with Dulcitol, D-Alanine, D-Galactose, D-Saccharic Acid, and L-Arabinose as carbon sources. The growth rates in LB medium, D-Mannose, D-Trehalose, D-Glucose, L-Aspartic-Acid, L-Proline, N-Acetyl-D-Glucosamine, and Succinic-Acid are presented in Table 4.5.1.

Table 4.5.1. *In silico* growth rates in different media

	***In silico* growth rate (h^{-1})**
***In silico* growth in LB medium**	79.6347
In silico* growth in Base medium	1.3070
***In silico* growth in D_Mannose**	0.4747
***In silico* growth in D_Trehalose**	0.4747
***In silico* growth in Glucose**	0.4747
***In silico* growth in L_Aspartic_Acid**	0.3569
***In silico* growth in L_Proline**	0.3714
***In silico* growth in N_Acetyl_D_Glucosamine**	0.8713
***In silico* growth in Succinic_Acid**	0.3569
***In silico* growth in Dulcitol**	no growth
***In silico* growth in D_Alanine**	no growth
***In silico* growth in D_Galactose**	no growth
***In silico* growth in D_Saccharic_Acid**	no growth
***In silico* growth in L_Arabinose**	no growth

* Based on the minimal media required for *in silico* growth in glucose
[c] Based on media in SEED Model database

The gene essentiality for *in silico* growth was compared to the predictions made using TRADIS. For each single gene deletion, the ability to grow is

assessed in the given media. The essentiality predictions for all the genes present in the model are presented in Table 4.5.2. Genes highlighted in blue are true positives, which were identified as essential both in LB medium simulations and in TRADIS. Genes highlighted in green are false positives (essential in LB medium simulations, not essential in TRADIS); in red are false negatives (not essential in LB medium simulations, essential in TRADIS); in yellow are true negatives (not essential in LB medium simulations and in TRADIS). The number of genes is presented in Table 4.52. We chose the LB medium simulations for comparison as the TRADIS experiments are performed in LB medium. The F score is calculated as 0.47. For established model organisms such as E. coli, about 90% of essential genes can be predicted using FBA [40]. Lower accuracy (60-80%) is achieved for other organisms [239,240]. Although *B. pseudomallei* FBA model presented here is not curated with any experimental data, essential genes are predicted with 79% accuracy. This attests to the power of the FBA model. The identification of essential genes and non-redundant metabolic pathways could aid to identify potential new drug targets.

Table 4.5.2 Comparison of the number of genes predicted to be essential / not essential in simulations of the genome-scale metabolic *B. pseudomallei* model versus experimental data obtained using TRADIS.

		Gene essentiality using TRADIS	
		Essential	Not essential
Gene essentiality predictions *in silico* growth in LB medium	Essential	113	156
	Not essential	98	815

4.6 Reduction of the genome-scale metabolic network

The availability of fully sequenced annotated genomes has facilitated the construction of genome-scale models for different organisms. The complexity and capacity of these models are expanding with new biological knowledge. High amounts of computational power can be required to analyse genome-scale metabolic models. As the size of the models increase, it becomes more difficult to apply certain stoichiometric methods (i.e. enumeration of elementary modes). Furthermore, capturing basic metabolic principles are not easily achievable when one has to interpret huge numbers of patterns arising from thousands of reactions and metabolites. Often the fundamental features of an organism's metabolic regulation are governed by core networks. Thus, it is useful to reduce the genome-scale models into smaller core models able to capture the capacity of the large models. This also simplifies experimental efforts of having to measure large sets of fluxes for validation purposes. Parameter dependent models, such as kinetic models, can only be reliably developed for small-scale

models. Network reduction allows focusing on the sub-modules of interest within the genome-scale network.

The *NetworkReducer* algorithm [241] offers an automated procedure to reduce the genome-scale metabolic constructions to meaningful core models preserving the capabilities of the large network. The algorithm accepts a set of reactions and metabolites together with a subset of protected metabolites and reactions. Often the reactions and metabolites of central metabolism are selected. A typical phenotype is to preserve biomass production. First, the feasibility of the protected functions is checked and the blocked reactions are removed. In the following network-pruning step, the non-protected parts of the network are deleted one by one if they are dispensable for the protected phenotypes. This can be followed by a compression step, which lumps the remaining reactions together to further reduce the network size.

The *NetworkReducer* algorithm was applied to the genome-scale model of *Burkholderia pseudomallei* K96243 SEED272560_3. By protecting biomass production on glucose uptake, the number of reactions was reduced from 1440 to 498 (*SEED272560_3_Pruned_1*, Table 4.6.1). The trehalose cycle was not preserved in the resulting network, *SEED272560_3_Pruned_1*, implying that it is possible for the reduced networks to lose non-protected functionalities. Nonetheless, it would be desirable to develop a model comprising the trehalose cycle to study the general structural properties of the cycle or to evaluate the synthesis of certain products or to discover various optimal states. The network reduction process was repeated by protecting trehalose uptake and utilization reactions in addition to central metabolism (biomass production on glucose uptake is kept as the protected phenotype). The resulting reduced network (*SEED272560_3_Pruned_2*) contained 564 reactions and includes the

trehalose pathways. Out of these 564 reactions, 499 are in common with the reactions of *SEED272560_3_Pruned_1*.

Table 4.6.1 The genome-scale and reduced models of *B. pseudomallei K96243 SEED272560_3*

	SEED272560_3	*SEED272560_3_Pruned_1*	*SEED272560_3_Pruned_2*
Metabolites/Compounds	1262	509	554
Reactions	1440	498	564
Growth rate (biomass production)	1.3070	1.3070	1.3070

4.7 Elementary flux mode analysis

The metabolic capacity of the system can be explored using Elementary Flux Mode analysis (EFM) [242]. EFM calculates the feasible routes through a network identifying steady state pathways. This allows the network to be decomposed into fundamental functional units. Within these identified functional routes, inhibition of one enzyme leads to failure of the steady-state flux in the network. Thus, it is possible to systematically analyse the essentiality of genes for certain capabilities of the network.

Enumeration of the full set of EFMs for large-scale models is highly computationally intensive as the algorithm iterates through a large number of combinatorial paths. While we were not able to enumerate the EFMs for the original *SEED272560_3* model, network reduction provided smaller-scale networks for which we could run the EFM analysis. The number of EFMs was considerably smaller in *SEED272560_3_Pruned_1* compared to *SEED272560_3_Pruned_2*. This could be due to the removal of redundant

paths in the *NetworkReducer* algorithm. Imposing the maintenance of the trehalose cycle reactions introduces a large number of combinatorial paths. While the number of EFMs for *SEED272560_3_Pruned_1* is sufficient small to interpret them biologically, *SEED272560_3_Pruned_2* allows us to explore the role of trehalase on the network capacity. When the trehalase gene was deleted, giving the model *SEED272560_3_Pruned_2 (ΔtreA)*, the number of EFMs decreases (Table 4.6.2). This could suggest a loss in metabolic capacity. The biological interpretation of the absent modes is to be devised. These modes can point to different ways of trehalose being utilised and fed into different parts of metabolism. The modes that are lost in the trehalase mutant can be essential for the metabolic network to operate in steady state for cell growth and maintenance. This would provide insight into the loss of virulence in the trehalase mutant.

Table 4.6.2. Elementary Flux modes of the genome-scale and reduced models of *Burkholderia pseudomallei K96243 SEED272560_3*

	Number of Elementary Flux Modes
SEED272560_3	Not enumerated
SEED272560_3_Pruned_2	1995515
SEED272560_3_Pruned_2 (ΔtreA)	1311832

Prior work has documented the significance of trehalase mutation in infection with *B. pseudomallei* [98]. Modelling approaches have been applied to decipher the mechanisms through which trehalose metabolism regulates the adaptation of yeast to changing conditions. However, these studies have only focused on trehalose-phosphate synthase mutations.

In this chapter, a detailed kinetic model of the trehalose cycle was studied to evaluate the effects on the metabolite concentrations of trehalase mutation. Transcriptomics data of *B. pseudomallei* was incorporated to simulate the dynamics of trehalose metabolism during infection. The interaction of trehalose cycle with glycolysis was studied for the trehalase mutation. Future studies should test the simulation results using laboratory experiments (i.e. evaluate the changes in metabolite concentrations using mass spectrometry). The trehalose metabolism in bacteria is not limited to the trehalose cycle (TPS/TPP pathway-trehalase). Future work should include additional interactions that *B. pseudomallei* trehalose may possess (such as trehalose synthase, Figure 4.7.1.).

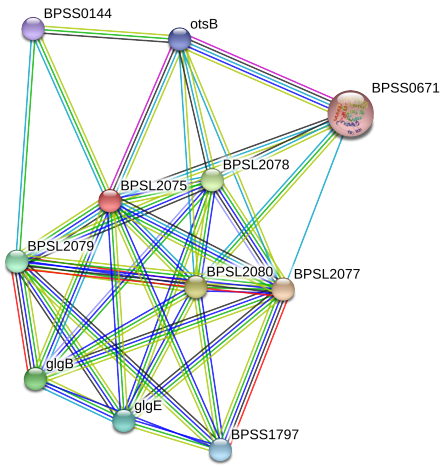


Figure 4.7.1. Network of proteins interacting with *B. pseudomallei* trehalase.

CHAPTER 5. CONCLUSIONS AND FUTURE WORK

5.1 Summary

It is crucial to provide completely new approaches to understanding the neglected *B. pseudomallei*, as this thesis has done, for the following reasons:

- *B. pseudomallei* is widely distributed in many regions worldwide and the number of melioidosis cases are much higher than expected [112].
- *B. pseudomallei* lives in the environment, soil and water. There are multiple infection routes; inhalation, ingestion, or skin inoculation.
- Melioidosis is greatly under-recognised, difficult to diagnose due to lack of disease-specific symptoms.
- *B. pseudomallei* is intrinsically antibiotic resistant and melioidosis has a high mortality rate.

Previous work has highlighted the dangers of overlooking this infection and calls for immediate control measures, but the understanding of *B. pseudomallei* interference and melioidosis is far from complete. Contributing to this understanding, this thesis explored a systems approach to expand the existing knowledge about the bacterium's ability to survive and infect host cells.

- A novel approach employed by this thesis was to investigate the role of host DNA methylation during *B. pseudomallei* infection. The findings presented in this thesis provided the first compelling evidence of epigenetic changes in the host occurring following *B. pseudomallei* infection.
- Equally importantly, this thesis employed mathematical modelling to elucidate the metabolic regulation of *B. pseudomallei*. The predictions

demonstrated in this thesis emphasise mechanisms of which the trehalose pathway may contribute to infection.

One particular strength of this thesis's methodology and outcomes is that such approaches can be applied to other diseases and pathogens. Due to its interdisciplinary nature, this thesis was able to demonstrate the benefits of combining biological experiments with mathematical analyses to achieve a more complete picture of disease. This will stimulate a wider understanding of the applications of systems biology to host-pathogen research and defence needs.

The main findings of this thesis are as follows. Chapter 2 established an in vitro infection model to quantify genome-wide patterns of DNA methylation in *B. pseudomallei* infected human macrophages. It offered the following discoveries:

- Reproducible infection induced differentially methylated positions (DMPs) were identified.
- DMPs were located in heavily, partially, and sparsely methylated regions.
- Most of the methylation changes in infected macrophages were losses in methylation rather than gains.
- There were five different differential methylation patterns: constant, early, late, transient, and oscillatory responses.
- A range of genes associated with the immune response to infection were differentially methylated, e.g. immune responses including cytokines and chemokines. DMPs were located in the vicinity of genes involved in inflammatory responses, actin regulation, nitric-oxide generation, T-cell responses, ubiquitination.

- qPCR analysis and comparisons with publically available transcription data identified genes that are both differentially methylated and differentially expressed.
- Comparison with the publically available DNA methylation changes during infection with other pathogens identified pathogen-specific and pathogen-common changes.

In Chapter 3, whole blood DNA methylation levels of patients diagnosed with septicemic melioidosis and healthy subjects were measured. Four major discoveries emerged:

- DNA methylation profiles of septicemic melioidosis patients were distinct from those of healthy individuals and individuals with type 2 diabetes mellitus.
- Differentially methylated positions in melioidosis patients were generally hypo-methylated compared to healthy subjects.
- There were also differences in DNA methylation on large genomic regions (DMRs).
- DMPs and DMRs map to genes associated with hyper-inflammatory and immunosuppressive responses attesting to their potential use as biomarkers.

Chapter 4 applied mathematical modelling approaches to investigate the role trehalose metabolism in virulence. These approaches can be summarised as:

- Trehalose pathway of *B. pseudomallei* was constructed using literature survey and databases. The genes/enzymes involved in *B. pseudomallei* trehalose pathway had high sequence identity to *E. coli* and *S. cerevisiae* genes/enzymes.

- While in yeast, glycolysis is affected in trehalose pathway mutant via regulation of the first step of glycolysis, no evidence was found for this mechanism in *B. pseudomallei*.
- Simulations of trehalase mutation in kinetic model of yeast and the model modified for *B. pseudomallei* showed similar responses (i.e. increase or decrease in the metabolite concentration, accumulation of trehalose).
- *B. pseudomallei* transcriptional changes were incorporated into the kinetic model.
- A single gene deletion analysis was performed on the genome-scale metabolic model of *B. pseudomallei*.
- The genome-scale metabolic model of *B. pseudomallei* was reduced to a core model and elementary flux modes were calculated.

The findings presented in Chapter 2 and Chapter 3 provide evidence for the role of DNA methylation in infection. The *in vitro* model established in Chapter 2 enables measuring DNA methylation changes in a controlled environment. As DNA methylation is greatly affected by environmental factors, the use of an *in vitro* model is beneficial in detecting differences that are solely due to infection. Elaborate study designs (i.e multiple time points) can be developed using *in vitro* models. However, *in vitro* studies do not always reflect the physiology of the human body. In fact, DNA methylation signatures identified in Chapter 2 are mostly distinct from the signatures identified in Chapter 3. The *in vivo* study established in Chapter 3 is beneficial in identifying biomarker panels for infection directly in patients. Not only these markers are biologically more relevant, DNA methylation signatures observed in the human blood can be effectively harnessed to develop diagnostic tests as it is a fairly easy procedure

to draw blood from patients. However, accurate results can only be obtained after validation in large cohorts. The combination of *in vitro* and *in vivo* studies is definitely beneficial in developing the most accurate biomarkers. Integration of *in silico* models can be central to study highly pathogenic bacteria such as *B. pseudomallei*. Chapter 4 demonstrated that mathematical modelling approaches could provide insight into bacterial metabolism. Exploiting these computational techniques is essential when wet laboratory experimentation is difficult due to the nature of the bacteria under study. Ultimately, integration of data from *in vitro*, *in vivo* and *in silico* models will lead to better understanding of complex disease.

5.2 Limitations of the study

As with all studies, the work presented in this thesis has a number of limitations to be considered. Although Chapter 2 measures the DNA methylation during the time course of infection, this is only during the early stages of infection. Only 1 sample of RNA could be collected for each group in the replication experiment presented in Chapter 2. As this limits the power to detect significant associations with gene expression, no analysis was made in the general transcriptome response of U937 cells following infection by *B. pseudomallei*.

Further, Infinium HumanMethylation450 BeadChip is the widely-used established technology for human methylomic analysis. Nonetheless, it should be taken into consideration that as it is a microarray-based technology it relies on predetermined probes to target methylation sites. Unlike bisulfite-sequencing technologies, Infinium HumanMethylation450 BeadChip (recently replace by Infinium MethylationEPIC Kit with greater coverage) can not be used to target

new regions of interest. This limits the coverage of the human methylome achieved in this analysis and it should be noted that there could be infection-induced differences in parts of the genome, which is not covered by the array.

The study design employed in Chapter 2 does not answer whether differential methylation is a consequence of the direct interaction of bacterial methyltransferases with host cell DNA, or the consequence of modification of host cell methyltransferase activity.

Further caveats are also required. While chapter 3 provides compelling evidence for the presence of DNA methylation differences in melioidosis patients, these differences should only be accepted as preliminary candidates for biomarkers as the number of samples assessed in this thesis is relatively low (i.e. discovery of clinically relevant biomarkers have been possible with sample sizes of hundreds [158,243,244]).

A number of factors, such as medications and comorbidities, may affect analyses that used blood samples [139,243,245]. While the presence of comorbidities were taken into account in Chapter 3, the extent of effects each disease contributes to DNA methylation could not be distinguished. No information was available about the medications used by the patients. Smoking can be an important confounding factor in DNA methylation studies [246]; in this thesis the smoking status of the subjects were not known. Such factors could complicate data interpretation.

There was no information on DNA methylation differences in sepsis caused by other pathogens. Therefore the analyses employed in this thesis could not distinguish whether the observed methylation differences were specific to septicemic melioidosis or common to sepsis caused by other infections.

Cell types also affect the conclusions drawn from the present research. This thesis compared genome-wide methylation differences between *B. pseudomallei* and *M. tuberculosis* [80] to identify pathogen-specific and pathogen-generic signatures. These studies used different cell types. As epigenetic processes exhibit cell type-specific patterns [247], this comparison has a limited power. Another factor complicating the direct comparison is that the studies used different technologies to measure DNA methylation and thus their coverage of the genome is different.

The work discussed in Chapter 4 suffers from a lack of experimental data on the *B. pseudomallei* trehalose pathway. Enzyme kinetics of model organisms was used to model *B. pseudomallei* pathways. The metabolite concentrations were simulated, yet the predictions could not be tested against experimental levels. As such, the *B. pseudomallei* genome-scale reconstruction model used in Chapter 4 is a draft model rather than a complete curated model (not fitted to experimental data). This greatly affects the power of predictions made by the downstream gene essentiality and elementary flux mode analyses.

5.3 Future work

This thesis provides evidence for the role of DNA methylation in infectious diseases and provides a framework for future studies to assess the relation between metabolism and virulence.

The presence of such DNA methylation differences associated with *B. pseudomallei* infection is an important new finding. Indeed, the work presented here offers exciting avenues for further research, and there are several ways that the findings of this thesis can be extended through new study..

Chapter 2 concentrated on understanding whether *B. pseudomallei* modulated DNA methylation in infected macrophages, rather than fully understanding the downstream consequences. A comprehensive temporal analysis of gene expression and proteome will enable to fully interpret the methylation data. In future, it is advisable that samples are taken at additional later time points to fully assess the time course of infection in macrophages. Further work will benefit from exploring whether the changes are pathogen-initiated or pathogen-mediated. To this end, dead bacterial cells (i.e. heat killed as shown in Marr *et al*. [81]) can be used as an additional control in the infection model in order to understand whether just the passive presence of the molecular components of the pathogen are sufficient to initiate host response or whether the pathogen actively elicits the response. An alternative approach might assess the similarity between the host- and pathogen DNA methyltransferases. None of the bacterial effectors identified to date is known actively to modulate DNA methylation. However, only a limited number of effectors have been identified to date, and the repertoire is likely to be much greater than the current list. A study of the similarity of the enzymes of the host and the pathogen can be used to infer whether bacteria are capable of directly methylating host DNA. If so, blocking the bacterial enzymes' function could be a further experiment to pursue.

The associations presented in Chapter 3 need to be validated in larger cohorts. Further work is planned to recruit more melioidosis patients to achieve statistically powerful results. Care will be taken to collect information on subjects that can be a confounding factor such as received treatments, smoking status and underlying diseases. Longitudinally sampled cohorts will be recruited to assess the dynamics of DNA methylation changes during the course of

melioidosis. Subjects can be grouped in days post admission to hospital to assess the methylation changes during acute infection. Furthermore, samples from chronic melioidosis patients can be assessed to decipher whether the markers identified in acute patients are conserved. It would certainly be valuable to evaluate DNA methylation changes during sepsis caused by other pathogens. In fact, assessing patients with ordinary infections (i.e. common flu) may inform whether these methylation changes are specific to highly pathogenic species or are an ordinary host response to infection.

Identifying methylation changes opens new opportunities to control infection. Currently, there are small molecules that manipulate epigenetic modifications: 5-Azacytidine and 5-aza-2'-deoxycitidine are the two most potent DNMT inhibitors that exhibit hypomethylating activity. However, these drugs act globally. Identifying and understanding specific methylation biomarkers will aid the development of specialized drugs in future.

The concluding remarks of Chapter 4 emphasise the need for experimental data to validate the mathematical models. Mass spectrometry should be used to measure whether the metabolite levels shown in the model reflect the physiological behaviour of the system. Enzyme kinetic assays can help improve models by providing the kinetic constants specific for *B. pseudomallei*. The improved models, then, will have predictive power to test hypothesis, which are not feasibly tested experimentally.

The trehalose metabolism in bacteria is not limited to the trehalose cycle (TPS/TPP pathway-trehalase). Future work should include any additional interactions *B. pseudomallei* trehalose may possess. Interestingly, a link between bacterial glycogen-, maltose-, and trehalose-metabolism has been

discovered [248]. It would be beneficial to model these interactions to provide a complete understanding of the dynamics of trehalose metabolism.

Advancing the genome-scale metabolic models of *B. pseudomallei* will provide a framework to integrate information from a multitude of sources. Transciptomics, proteomics, metabolomics and epigenetics data of the host and pathogen can be incorporated into such models to achieve a holistic view of the host-pathogen interaction.

Bibliography

1. WHO (2003) WHO World Health Organisation. WHO Defin Heal 1948. Available: http://www.who.int/about/definition/en/print.html.
2. Ventola CL (2015) The antibiotic resistance crisis. P T A peer-reviewed J Formul Manag 40: 277–283. doi:Article.
3. Oyston P, Robinson K (2012) The current challenges for vaccine development. J Med Microbiol 61: 889–894. doi:10.1099/jmm.0.039180-0.
4. Bhavsar AP, Guttman J a, Finlay BB (2007) Manipulation of host-cell pathways by bacterial pathogens. Nature 449: 827–834. doi:10.1038/nature06247.
5. Ge SX (2011) Large-scale analysis of expression signatures reveals hidden links among diverse cellular processes. BMC Syst Biol 5: 87. doi:10.1186/1752-0509-5-87.
6. Eckhart AD, Beebe K, Milburn M (2012) Metabolomics as a Key Integrator for “Omic” Advancement of Personalized Medicine and Future Therapies. Clin Transl Sci 5: 285–288. doi:10.1111/j.1752-8062.2011.00388.x.
7. Waterer GW (2012) Community-acquired pneumonia: Genomics, epigenomics, transcriptomics, proteomics, and metabolomics. Semin Respir Crit Care Med 33: 257–265. doi:10.1055/s-0032-1315637.
8. Valdes AM, Glass D, Spector TD (2013) Omics technologies and the study of human ageing. Nat Rev Genet 14: 601–607. doi:10.1038/nrg3553.
9. Bains RK (2014) Human infectious diseases in the genomics era: where do we go from here? Genome Biol 15: 529. doi:10.1186/s13059-014-0529-5.

10. Gire SK, Goba A, Andersen KG, Sealfon RSG, Park DJ, et al. (2014) Genomic surveillance elucidates Ebola virus origin and transmission during the 2014 outbreak. Science 345: 1369–1372. doi:10.1126/science.1259657.

11. Lei F, Shi W (2011) Prospective of Genomics in Revealing Transmission, Reassortment and Evolution of Wildlife-Borne Avian Influenza A (H5N1) Viruses. Curr Genomics 12: 466–474. doi:10.2174/138920211797904052.

12. Foxman B (2012) Molecular Tools and Infectious Disease Epidemiology. 9-21 p. doi:10.1016/B978-0-12-374133-2.00002-2.

13. Schulze S, Henkel SG, Driesch D, Guthke R, Linde J (2015) Computational prediction of molecular pathogen-host interactions based on dual transcriptome data. Front Microbiol 6. doi:10.3389/fmicb.2015.00065.

14. Zhou H, Jin J, Wong L (2013) Progress in computational studies of host-pathogen interactions. J Bioinform Comput Biol 11: 1230001. doi:10.1142/S0219720012300018.

15. Wang Z, Gerstein M, Snyder M (2009) RNA-Seq: a revolutionary tool for transcriptomics. Nat Rev Genet 10: 57–63. doi:10.1038/nrg2484.

16. Kim Y-M, Schmidt BJ, Kidwai AS, Jones MB, Deatherage Kaiser BL, et al. (2013) Salmonella modulates metabolism during growth under conditions that induce expression of virulence genes. Mol Biosyst 9: 1522. doi:10.1039/c3mb25598k.

17. Mendum TA, Wu H, Kierzek AM, Stewart GR (2015) Lipid metabolism and Type VII secretion systems dominate the genome scale virulence profile of *Mycobacterium tuberculosis* in human dendritic cells. BMC

Genomics 16: 372. doi:10.1186/s12864-015-1569-2.

18. Navid A, Almaas E (2012) Genome-level transcription data of Yersinia pestis analyzed with a new metabolic constraint-based approach. BMC Syst Biol 6: 150. doi:10.1186/1752-0509-6-150.

19. Bumann D (2009) System-level analysis of Salmonella metabolism during infection. Curr Opin Microbiol 12: 559–567. doi:10.1016/j.mib.2009.08.004.

20. Berger SL, Kouzarides T, Shiekhattar R, Shilatifard A (2009) An operational definition of epigenetics. Genes Dev 23: 781–783. doi:10.1101/gad.1787609.

21. Kouzarides T (2007) Chromatin modifications and their function. Cell 128: 693–705. doi:10.1016/j.cell.2007.02.005.

22. Quackenbush J (2001) Computational analysis of microarray data. Nat Rev Genet 2: 418–427. doi:10.1038/35076576.

23. D'Haeseleer P (2005) How does gene expression clustering work? Nat Biotechnol 23: 1499–1501. doi:10.1038/nbt1205-1499.

24. Ringnér M, Ringner M (2008) What is principal component analysis? Nat Biotechnol 26: 303–304. doi:10.1038/nbt0308-303.

25. Gene Ontology Consortium (2015) Gene Ontology Consortium: going forward. Nucleic Acids Res 43: D1049–D1056. doi:10.1093/nar/gku1179.

26. Ogata H, Goto S, Sato K, Fujibuchi W, Bono H, et al. (1999) KEGG: Kyoto encyclopedia of genes and genomes. Nucleic Acids Res 27: 29–34. doi:10.1093/nar/27.1.29.

27. Kitano H (2002) Computational systems biology. Nature 420: 206–210. doi:10.1038/nature01254.

28. Kitano H (2002) Systems biology: a brief overview. Science 295: 1662–

1664. doi:10.1126/science.1069492.

29. Durmus S, Sakir T, Ozgur A, Guthke R (2015) A review on computational systems biology of pathogen-host interactions. Front Microbiol 6. doi:10.3389/fmicb.2015.00235.

30. Schleicher J, Conrad T, Gustafsson M, Cedersund G, Guthke R, et al. (2016) Facing the challenges of multiscale modelling of bacterial and fungal pathogen-host interactions. Brief Funct Genomics: elv064 – . doi:10.1093/bfgp/elv064.

31. Szallasi Z (2006) System Modeling in Cellular Biology. 452 p. doi:10.7551/mitpress/9780262195485.001.0001.

32. Teusink B, Passarge J, Reijenga CA, Esgalhado E, Weijden CC Van Der, et al. (2000) Can yeast glycolysis be understood in terms of in vitro kinetics of the constituent enzymes ? Testing biochemistry. 5329.

33. Chassagnole C, Noisommit-Rizzi N, Schmid JW, Mauch K, Reuss M (2002) Dynamic modeling of the central carbon metabolism of *Escherichia coli*. Biotechnol Bioeng 79: 53–73. doi:10.1002/bit.10288.

34. Kotte O, Zaugg JB, Heinemann M (2010) Bacterial adaptation through distributed sensing of metabolic fluxes. Mol Syst Biol 6: 355. doi:10.1038/msb.2010.10.

35. Kotte O, Volkmer B, Radzikowski JL, Heinemann M (2014) Phenotypic bistability in *Escherichia coli*'s central carbon metabolism. Mol Syst Biol 10: 736. doi:10.15252/msb.20135022.

36. Feist AM, Herrgård MJ, Thiele I, Reed JL, Palsson BØ (2009) Reconstruction of biochemical networks in microorganisms. Nat Rev Microbiol 7: 129–143. doi:10.1038/nrmicro1949.

37. Oberhardt M a, Palsson BØ, Papin J a (2009) Applications of genome-

scale metabolic reconstructions. Mol Syst Biol 5: 320. doi:10.1038/msb.2009.77.

38. Bordbar A, Palsson BO (2012) Using the reconstructed genome-scale human metabolic network to study physiology and pathology. J Intern Med 271: 131–141. doi:10.1111/j.1365-2796.2011.02494.x.

39. Orth JD, Thiele I, Palsson BØ (2010) What is Flux Balance Analysis ? Nat Biotechnol 28: 245–248. doi:10.1038/nbt.1614.

40. Feist AM, Henry CS, Reed JL, Krummenacker M, Joyce AR, et al. (2007) A genome-scale metabolic reconstruction for *Escherichia coli K-12* MG1655 that accounts for 1260 ORFs and thermodynamic information. Mol Syst Biol 3: 121. doi:10.1038/msb4100155.

41. Bordbar A, Lewis NE, Schellenberger J, Palsson Bernhard Øand Jamshidi N, Palsson BØ, et al. (2010) Insight into human alveolar macrophage and *M. tuberculosis* interactions via metabolic reconstructions. Mol Syst Biol 6: 422. doi:10.1038/msb.2010.68.

42. Raghunathan A, Reed J, Shin S, Palsson B, Daefler S (2009) Constraint-based analysis of metabolic capacity of *Salmonella typhimurium* during host-pathogen interaction. BMC Syst Biol 3: 38. doi:10.1186/1752-0509-3-38.

43. Raghunathan A, Shin S, Daefler S (2010) Systems approach to investigating host-pathogen interactions in infections with the biothreat agent *Francisella.* Constraints-based model of *Francisella tularensis.* BMC Syst Biol 4: 118. doi:10.1186/1752-0509-4-118.

44. Bird A (2007) Perceptions of epigenetics. Nature 447: 396–398. doi:10.1038/nature05913.

45. Bird AP (1986) CpG-rich islands and the function of DNA methylation.

Nature 321: 209–213. doi:10.1038/321209a0.

46. Jenuwein T, Allis CD (2001) Translating the histone code. Science 293: 1074–1080. doi:10.1126/science.1063127.

47. Ziller MJ, Gu H, Müller F, Donaghey J, Tsai LT-Y, et al. (2013) Charting a dynamic DNA methylation landscape of the human genome. Nature 500: 477–481. doi:10.1038/nature12433.

48. Gutierrez-Arcelus M, Ongen H, Lappalainen T, Montgomery SB, Buil A, et al. (2015) Tissue-Specific Effects of Genetic and Epigenetic Variation on Gene Regulation and Splicing. PLoS Genet 11. doi:10.1371/journal.pgen.1004958.

49. Slatkin M (2009) Epigenetic inheritance and the missing heritability problem. Genetics 182: 845–850. doi:10.1534/genetics.109.102798.

50. Hackett JA, Sengupta R, Zylicz JJ, Murakami K, Lee C, et al. (2013) Germline DNA demethylation dynamics and imprint erasure through 5-hydroxymethylcytosine. Science 339: 448–452. doi:10.1126/science.1229277.

51. Khavari DA, Sen GL, Rinn JL (2010) DNA methylation and epigenetic control of cellular differentiation. Cell Cycle 9: 3880–3883. doi:10.4161/cc.9.19.13385.

52. Bollati V, Baccarelli A (2010) Environmental epigenetics. Heredity (Edinb) 105: 105–112. doi:10.1038/hdy.2010.2.

53. Meaney MJ (2013) Epigenetics and the Environmental Regulation of the Genome and Its Function. Evolution, Early Experience and Human Development: From Research to Practice and Policy. doi:10.1093/acprof:oso/9780199755059.003.0006.

54. Bernstein BE, Meissner A, Lander ES (2007) The Mammalian

Epigenome. Cell 128: 669–681. doi:10.1016/j.cell.2007.01.033.

55. Ozanne SE, Constância M (2007) Mechanisms of Disease: the developmental origins of disease and the role of the epigenotype. Nat Rev Endocrinol 3: 539–546. doi:10.1038/ncpendmet0531.

56. Laird PW (2003) The power and the promise of DNA methylation markers. Nat Rev Cancer 3: 253–266. doi:10.1038/nrc1045.

57. Mulero-Navarro S, Esteller M (2008) Epigenetic biomarkers for human cancer: The time is now. Crit Rev Oncol Hematol 68: 1–11. doi:10.1016/j.critrevonc.2008.03.001.

58. Luger K, Mäder a W, Richmond RK, Sargent DF, Richmond TJ (1997) Crystal structure of the nucleosome core particle at 2.8 A resolution. Nature 389: 251–260. doi:10.1038/38444.

59. Mattick JS, Makunin I V. (2006) Non-coding RNA. Hum Mol Genet 15 Spec No. doi:10.1093/hmg/ddl046.

60. Klose RJ, Bird AP (2006) Genomic DNA methylation: The mark and its mediators. Trends Biochem Sci 31: 89–97. doi:10.1016/j.tibs.2005.12.008.

61. Deaton AM, Bird A (2011) CpG islands and the regulation of transcription. Genes Dev 25: 1010–1022.

62. Zakhari S (2013) Alcohol metabolism and epigenetics changes. Alcohol Res 35: 6–16.

63. Chen T, Ueda Y, Dodge JE, Wang Z, Li E (2003) Establishment and maintenance of genomic methylation patterns in mouse embryonic stem cells by Dnmt3a and Dnmt3b. Mol Cell Biol 23: 5594–5605. doi:10.1128/MCB.23.16.5594.

64. Tahiliani M, Koh KP, Shen Y, Pastor WA, Bandukwala H, et al. (2009)

Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. Science (80-) 324: 930–935. doi:10.1126/science.1170116.

65. Zhang H, Zhu JK (2012) Active DNA demethylation in plants and animals. Cold Spring Harb Symp Quant Biol 77: 161–173. doi:10.1101/sqb.2012.77.014936.

66. Kohli RM, Zhang Y (2013) TET enzymes, TDG and the dynamics of DNA demethylation. Nature 502: 472–479. doi:10.1038/nature12750.

67. Wu H, Zhang Y (2014) Reversing DNA methylation: Mechanisms, genomics, and biological functions. Cell 156: 45–68. doi:10.1016/j.cell.2013.12.019.

68. Watt F, Molloy PL (1988) Cytosine methylation prevents binding to DNA of a HeLa cell transcription factor required for optimal expression of the adenovirus major late promoter. Genes Dev 2: 1136–1143. doi:10.1101/gad.2.9.1136.

69. Boyes J, Bird A (1991) DNA methylation inhibits transcription indirectly via a methyl-CpG binding protein. Cell 64: 1123–1134. doi:10.1016/0092-8674(91)90267-3.

70. Lorincz MC, Dickerson DR, Schmitt M, Groudine M (2004) Intragenic DNA methylation alters chromatin structure and elongation efficiency in mammalian cells. Nat Struct Mol Biol 11: 1068–1075. doi:10.1038/nsmb840.

71. Heard E, Disteche CM (2006) Dosage compensation in mammals: Fine-tuning the expression of the X chromosome. Genes Dev 20: 1848–1867. doi:10.1101/gad.1422906.

72. Li E, Beard C, Jaenisch R (1994) Role for DNA methylation in genomic

imprinting. Cell 366: 362–365.

73. Meng H, Meng H, Cao Y, Qin J, Song X, et al. (2015) DNA methylation, its mediators and genome integrity. Int J Biol Sci 11: 604–617. doi:10.7150/ijbs.11218.

74. Bracken AP, Helin K (2009) Polycomb group proteins: navigators of lineage pathways led astray in cancer. Nat Rev Cancer 9: 773–784.

75. Esteller M (2007) Cancer epigenomics: DNA methylomes and histone-modification maps. Nat Rev Genet 8: 286–298.

76. Jones PA, Baylin SB (2007) The epigenomics of cancer. Cell 128: 683–692.

77. Robertson KD (2005) DNA methylation and human disease. Nat Rev Genet 6: 597–610.

78. Paschos K, Allday MJ (2010) Epigenetic reprogramming of host genes in viral and microbial pathogenesis. Trends Microbiol 18: 439–447. doi:10.1016/j.tim.2010.07.003.

79. Gómez-Díaz E, Jordà M, Peinado MA, Rivero A (2012) Epigenetics of host-pathogen interactions: the road ahead and the road behind. PLoS Pathog 8: e1003007. doi:10.1371/journal.ppat.1003007.

80. Pacis A, Tailleux L, Morin AM, Lambourne J, MacIsaac JL, et al. (2015) Bacterial infection remodels the DNA methylation landscape of human dendritic cells. Genome Res 25: 1801–1811. doi:10.1101/gr.192005.115.

81. Marr AK, MacIsaac JL, Jiang R, Airo AM, Kobor MS, et al. (2014) Leishmania donovani Infection Causes Distinct Epigenetic DNA Methylation Changes in Host Macrophages. PLoS Pathog 10: e1004419. doi:10.1371/journal.ppat.1004419.

82. Park S, Yoo EJ, Cho N, Kim N, Kang GH (2009) Comparison of CpG

island hypermethylation and repetitive DNA hypomethylation in premalignant stages of gastric cancer , stratified for *Helicobacter pylori* infection: 410–416. doi:10.1002/path.

83. Qian X, Huang C, Cho CH, Hui WM, Rashid A, et al. (2008) E-cadherin promoter hypermethylation induced by interleukin-1beta treatment or H. pylori infection in human gastric cancer cell lines. Cancer Lett 263: 107–113. doi:10.1016/j.canlet.2007.12.023.

84. Hmadcha a, Bedoya FJ, Sobrino F, Pintado E (1999) Methylation-dependent gene silencing induced by interleukin 1beta via nitric oxide production. J Exp Med 190: 1595–1604.

85. Niwa T, Tsukamoto T, Toyoda T, Mori A, Tanaka H, et al. (2010) Inflammatory processes triggered by *Helicobacter pylori* infection cause aberrant DNA methylation in gastric epithelial cells. Cancer Res 70: 1430–1440. doi:10.1158/0008-5472.CAN-09-2755.

86. Takahashi K, Sugi Y, Nakano K, Tsuda M, Kurihara K, et al. (2011) Epigenetic control of the host gene by commensal bacteria in large intestinal epithelial cells. J Biol Chem 286: 35755–35762. doi:10.1074/jbc.M111.271007.

87. Brown SA, Palmer KL, Whiteley M (2008) Revisiting the host as a growth medium. Nat Rev Microbiol 6: 657–666. doi:10.1038/nrmicro1955.

88. Rohmer L, Hocquet D, Miller SI (2011) Are pathogenic bacteria just looking for food? Metabolism and microbial pathogenesis. Trends Microbiol 19: 341–348. doi:10.1016/j.tim.2011.04.003.

89. Toft C, Andersson SGE (2010) Evolutionary microbial genomics: insights into bacterial host adaptation. Nat Rev Genet 11: 465–475. doi:10.1038/nrg2798.

90. Martínez JL, Rojo F (2011) Metabolic regulation of antibiotic resistance. FEMS Microbiol Rev 35: 768–789. doi:10.1111/j.1574-6976.2011.00282.x.

91. Tournu H, Fiori A, Van Dijck P (2013) Relevance of Trehalose in Pathogenicity: Some General Rules, Yet Many Exceptions. PLoS Pathog 9: 6–9. doi:10.1371/journal.ppat.1003447.

92. Elbein AD, Pan YT, Pastuszak I, Carroll D (2003) New insights on trehalose: a multifunctional molecule. Glycobiology 13: 17R – 27R. doi:10.1093/glycob/cwg047.

93. Iturriaga G, Suárez R, Nova-Franco B (2009) Trehalose metabolism: From osmoprotection to signaling. Int J Mol Sci 10: 3793–3810. doi:10.3390/ijms10093793.

94. Ponnu J, Wahl V, Schmid M (2011) Trehalose-6-phosphate: connecting plant metabolism and development. Front Plant Sci 2: 70. doi:10.3389/fpls.2011.00070.

95. Kalscheuer R, Weinrick B, Veeraraghavan U, Besra GS, Jacobs WR (2010) Trehalose-recycling ABC transporter LpqY-SugA-SugB-SugC is essential for virulence of Mycobacterium tuberculosis. Proc Natl Acad Sci U S A 107: 21761–21766. doi:10.1073/pnas.1014642108.

96. Botts MR, Huang M, Borchardt RK, Hull CM, Hull CM (2014) Developmental cell fate and virulence are linked to trehalose homeostasis in Cryptococcus neoformans. Eukaryot Cell 13: 1158–1168. doi:10.1128/EC.00152-14.

97. Ruhal R, Kataria R, Choudhury B (2013) Trends in bacterial trehalose metabolism and significant nodes of metabolic pathway in the direction of trehalose accumulation. Microb Biotechnol 6: 493–502. doi:10.1111/1751-

7915.12029.

98. Vanaporn M, Sarkar-Tyson M, Kovacs-Simon A, Ireland PM, Pumirat P, et al. (2016) Trehalase plays a role in macrophage colonization and virulence of Burkholderia pseudomallei in insect and mammalian hosts. Http://DxDoiOrg/101080/2150559420161199316 5594. doi:10.1080/21505594.2016.1199316.

99. Benaroudj N, Lee DH, Goldberg AL (2001) Trehalose Accumulation during Cellular Stress Protects Cells and Cellular Proteins from Damage by Oxygen Radicals. J Biol Chem 276: 24261–24267. doi:10.1074/jbc.M101487200.

100. Zaragoza O, Blazquez MA, Gancedo C (1998) Disruption of the Candida albicans TPS1 gene encoding trehalose-6- phosphate synthase impairs formation of hyphae and decreases infectivity. J Bacteriol 180: 3809–3815.

101. Foster AJ, Jenkinson JM, Talbot NJ (2003) Trehalose synthesis and metabolism are required at different stages of plant infection by Magnaporthe grisea. EMBO J 22: 225–235. doi:10.1093/emboj/cdg018.

102. Lowe RGT, Lord M, Rybak K, Trengove RD, Oliver RP, et al. (2009) Trehalose biosynthesis is involved in sporulation of Stagonospora nodorum. Fungal Genet Biol 46: 381–389. doi:10.1016/j.fgb.2009.02.002.

103. Petzold EW, Himmelreich U, Mylonakis E, Rude T, Toffaletti D, et al. (2006) Characterization and regulation of the trehalose synthesis pathway and its importance in the pathogenicity of *Cryptococcus neoformans*. Infect Immun 74: 5877–5887. doi:10.1128/IAI.00624-06.

104. Ngamskulrungroj P, Price J, Sorrell T, Perfect JR, Meyer W (2011) *Cryptococcus gattii* virulence composite: Candidate genes revealed by

microarray analysis of high and less virulent Vancouver Island outbreak strains. PLoS One 6. doi:10.1371/journal.pone.0016076.

105. Freeman BC, Chen C, Beattie GA (2010) Identification of the trehalose biosynthetic loci of Pseudomonas syringae and their contribution to fitness in the phyllosphere. Environ Microbiol 12: 1486–1497. doi:10.1111/j.1462-2920.2010.02171.x.

106. Avonce N, Mendoza-Vargas A, Morett E, Iturriaga G (2006) Insights on the evolution of trehalose biosynthesis. BMC Evol Biol 6: 109. doi:10.1186/1471-2148-6-109.

107. Sasai-Takedatsu M, Taketani S, Nagata N, Furukawa T, Tokunaga R, et al. (1996) Human trehalase: characterization, localization, and its increase in urine by renal proximal tubular damage. Nephron 73: 179–185.

108. Eleutherio E, Panek A, De Mesquita JF, Trevisol E, Magalhães R (2014) Revisiting yeast trehalose metabolism. Curr Genet 61: 263–274. doi:10.1007/s00294-014-0450-1.

109. Empadinhas N (2014) The molecular biology of mycobacterial trehalose in the quest for advanced tuberculosis therapies. Microbiology 160: 1547–1570. doi:10.1099/mic.0.075895-0.

110. Wiersinga WJ, van der Poll T, White NJ, Day NP, Peacock SJ (2006) Melioidosis: insights into the pathogenicity of *Burkholderia pseudomallei.* Nat Rev Microbiol 4: 272–282. doi:10.1038/nrmicro1385.

111. Cheng AC, Currie BJ (2005) Melioidosis: Epidemiology, pathophysiology, and management. Clin Microbiol Rev 18: 383–416. doi:10.1128/CMR.18.2.383-416.2005.

112. Limmathurotsakul D, Golding N, Dance DAB, Messina JP, Pigott DM, et

al. (2016) Predicted global distribution of Burkholderia pseudomallei and burden of melioidosis. Nat Microbiol 1: 15008.

113. Hamid S. Jafari MD, Bradley A. Perkins MD, Jay D. Wenger MD (1996) Centers for Disease Control and Prevention (CDC). J Am Med Assoc 46: 955–958.

114. Meumann EM, Cheng AC, Ward L, Currie BJ (2012) Clinical features and epidemiology of melioidosis pneumonia: Results from a 21-year study and review of the literature. Clin Infect Dis 54: 362–369. doi:10.1093/cid/cir808.

115. Currie BJ, Fisher DA, Anstey NM, Jacups SP (2000) Melioidosis: acute and chronic disease, relapse and re-activation. Trans R Soc Trop Med Hyg 94: 301–304. doi:10.1016/S0035-9203(00)90333-X.

116. Suputtamongkol Y, Chaowagul W, Chetchotisakd P, Lertpatanasuwun N, Intaranongpai S, et al. (1999) Risk factors for melioidosis and bacteremic melioidosis. Clin Infect Dis 29: 408–413. doi:10.1086/520223.

117. Redondo MC, Gómez M, Landaeta ME, Ríos H, Khalil R, et al. (2011) Melioidosis presenting as sepsis syndrome: a case report. Int J Infect Dis 15: e217–e218. doi:10.1016/j.ijid.2010.11.009.

118. Wuthiekanun V, Smith MD, White NJ (1995) Survival of *Burkholderia pseudomallei* in the absence of nutrients. Trans R Soc Trop Med Hyg 89: 491.

119. Dejsirilert S, Kondo E, Chiewsilp D, Kanai K (1991) Growth and survival of *Pseudomonas pseudomallei* in acidic environments. Jpn J Med Sci Biol 44: 63–74.

120. Chen YS, Chen SC, Kao CM, Chen YL (2003) Effects of soil pH, temperature and water content on the growth of *Burkholderia*

pseudomallei. Folia Microbiol (Praha) 48: 253–256.

121. Holden MTG, Titball RW, Peacock SJ, Cerdeño-Tárraga AM, Atkins T, et al. (2004) Genomic plasticity of the causative agent of melioidosis, *Burkholderia pseudomallei.* Proc Natl Acad Sci U S A 101: 14240–14245. doi:10.1073/pnas.0403302101.

122. Chin C-Y, Monack DM, Nathan S (2010) Genome wide transcriptome profiling of a murine acute melioidosis model reveals new insights into how *Burkholderia pseudomallei* overcomes host innate immunity. BMC Genomics 11: 672. doi:10.1186/1471-2164-11-672.

123. Suparak S, Muangsombut V, Riyapa D, Stevens JM, Stevens MP, et al. (2011) *Burkholderia pseudomallei*-induced cell fusion in U937 macrophages can be inhibited by monoclonal antibodies against host cell surface molecules. Microbes Infect 13: 1006–1011. doi:10.1016/j.micinf.2011.06.007.

124. Pankla R, Buddhisa S, Berry M, Blankenship DM, Bancroft GJ, et al. (2009) Genomic transcriptional profiling identifies a candidate blood biomarker signature for the diagnosis of septicemic melioidosis. Genome Biol 10: R127.

125. Allwood EM, Devenish RJ, Prescott M, Adler B, Boyce JD (2011) Strategies for intracellular survival of *Burkholderia pseudomallei*. Front Microbiol 2. doi:10.3389/fmicb.2011.00170.

126. Chieng S, Carreto L, Nathan S (2012) *Burkholderia pseudomallei* transcriptional adaptation in macrophages. BMC Genomics 13: 328. doi:10.1186/1471-2164-13-328.

127. Brown NF, Boddey JA, Flegg CP, Beacham IR (2002) Adherence of *Burkholderia pseudomallei* cells to cultured human epithelial cell lines is

regulated by growth temperature. Infect Immun 70: 974–980.

128. Utaisincharoen P, Anuntagool N, Arjcharoen S, Lengwehasatit I, Limposuwan K, et al. (2004) *Burkholderia pseudomallei* stimulates low interleukin-8 production in the human lung epithelial cell line A549. Clin Exp Immunol 138: 61–65. doi:10.1111/j.1365-2249.2004.02601.x.

129. Balder R, Lipski S, Lazarus JJ, Grose W, Wooten RM, et al. (2010) Identification of *Burkholderia mallei* and *Burkholderia pseudomallei* adhesins for human respiratory epithelial cells. BMC Microbiol 10: 250. doi:10.1186/1471-2180-10-250.

130. Lehmann MH (1998) Recombinant human granulocyte-macrophage colony-stimulating factor triggers interleukin-10 expression in the monocytic cell line U937. Mol Immunol 35: 479–485.

131. Wand ME, Müller CM, Titball RW, Michell SL (2011) Macrophage and *Galleria mellonella* infection models reflect the virulence of naturally occurring isolates of *B. pseudomallei*, *B. thailandensis* and *B. oklahomensis*. BMC Microbiol 11: 11. doi:10.1186/1471-2180-11-11.

132. Wang RYH, Gehrke CW, Ehrlich M (1980) Comparison of bisulfite modification of 5-methyldeoxycytidine and deoxycytidine residues. Nucleic Acids Res 8: 4777–4790. doi:10.1093/nar/8.20.4777.

133. Wang T, Guan W, Lin J, Boutaoui N, Canino G, et al. (2015) A systematic study of normalization methods for Infinium 450K methylation data using whole-genome bisulfite sequencing data. Epigenetics 10: 662–669. doi:10.1080/15592294.2015.1057384.

134. Pidsley R, Zotenko E, Peters TJ, Lawrence MG, Risbridger GP, et al. (2016) Critical evaluation of the Illumina MethylationEPIC BeadChip microarray for whole-genome DNA methylation profiling. Genome Biol 17:

208. doi:10.1186/s13059-016-1066-1.

135. Bibikova M, Lin Z, Zhou L, Chudin E, Garcia EW, et al. (2006) High-throughput DNA methylation profiling using universal bead arrays. Genome Res 16: 383–393. doi:10.1101/gr.4410706.

136. Pidsley R, Y Wong CC, Volta M, Lunnon K, Mill J, et al. (2013) A data-driven approach to preprocessing Illumina 450K methylation array data. BMC Genomics 14: 293. doi:10.1186/1471-2164-14-293.

137. Bibikova M, Barnes B, Tsan C, Ho V, Klotzle B, et al. (2011) High density DNA methylation array with single CpG site resolution. Genomics 98: 288–295. doi:10.1016/j.ygeno.2011.07.007.

138. Price ME, Cotton AM, Lam LL, Farré P, Emberly E, et al. (2013) Additional annotation enhances potential for biologically-relevant analysis of the Illumina Infinium HumanMethylation450 BeadChip array. Epigenetics Chromatin 6: 4. doi:10.1186/1756-8935-6-4.

139. Mill J, Heijmans BT (2013) From promises to practical strategies in epigenetic epidemiology. Nat Rev Genet 14: 585–594. doi:10.1038/nrg3405.

140. McLean CY, Bristor D, Hiller M, Clarke SL, Schaar BT, et al. (2010) GREAT improves functional interpretation of cis-regulatory regions. Nat Biotechnol 28: 495–501. doi:10.1038/nbt.1630.

141. Fisher RA (1922) On the Interpretation of χ^2 from Contingency Tables, and the Calculation of P. J R Stat Soc 85: 87–94. doi:10.2307/2340521.

142. Young MD, Wakefield MJ, Smyth GK, Oshlack A (2010) Gene ontology analysis for RNA-seq: accounting for selection bias. Genome Biol 11: R14. doi:10.1186/gb-2010-11-2-r14.

143. Supek F, Bosnjak M, Skunca N, Smuc T (2011) Revigo summarizes and visualizes long lists of gene ontology terms. PLoS One 6. doi:10.1371/journal.pone.0021800.

144. Depke M, Breitbach K, Dinh Hoang Dang K, Brinkmann L, Salazar MG, et al. (2014) Bone marrow-derived macrophages from BALB/c and C57BL/6 mice fundamentally differ in their respiratory chain complex proteins, lysosomal enzymes and components of antioxidant stress systems. J Proteomics 103: 72–86. doi:10.1016/j.jprot.2014.03.027.

145. Radpour R, Sikora M, Grussenmeyer T, Kohler C, Barekati Z, et al. (2009) Simultaneous Isolation of DNA , RNA , and Proteins for Genetic , Epigenetic , Transcriptomic , and Proteomic Analysis research articles: 5264–5274.

146. Noble WS (2009) How does multiple testing correction work? Nat Biotechnol 27: 1135–1137. doi:10.1038/nbt1209-1135.

147. Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc B 57: 289–300. doi:10.2307/2346101.

148. Storey JD (2003) The positive false discovery rate: A Bayesian interpretation and the q-value. Ann Stat 31: 2013–2035. doi:10.1214/aos/1074290335.

149. Storey JD, Tibshirani R (2003) Statistical significance for genomewide studies. Proc Natl Acad Sci U S A 100: 9440–9445. doi:10.1073/pnas.1530509100.

150. Sham PC, Purcell SM (2014) Statistical power and significance testing in large-scale genetic studies. Nat Rev Genet 15: 335–346. doi:10.1038/nrg3706.

151. Dalman MR, Deeter A, Nimishakavi G, Duan Z-H (2012) Fold change and p-value cutoffs significantly alter microarray interpretations. file:///home/sbagew/PhD/workspace/publications/Co-expressionNetworks/Afroza/Fold_change/Gene_Category/GSE19677/ddq018.pdf 13: S11. doi:10.1186/1471-2105-13-S2-S11.

152. Conejero L, Potempa K, Graham CM, Spink N, Blankley S, et al. (2015) The blood transcriptome of experimental melioidosis reflects disease severity and shows considerable similarity with the human disease. J Immunol 195: 3248–3261. doi:10.4049/jimmunol.1500641.

153. Jaenisch R, Bird A (2003) Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. Nat Genet 33 Suppl: 245–254. doi:10.1038/ng1089.

154. Jones PA (2012) Functions of DNA methylation: islands, start sites, gene bodies and beyond. Nat Rev Genet 13: 484–492. doi:10.1038/nrg3230.

155. Nichols D, Chmiel J, Berger M (2008) Chronic inflammation in the cystic fibrosis lung: Alterations in inter- and intracellular signaling. Clin Rev Allergy Immunol 34: 146–162. doi:10.1007/s12016-007-8039-9.

156. Chiang C-Y, Ulrich RL, Ulrich MP, Eaton B, Ojeda JF, et al. (2015) Characterization of the murine macrophage response to infection with virulent and avirulent *Burkholderia* species. BMC Microbiol 15: 259. doi:10.1186/s12866-015-0593-3.

157. Alwatban A, Everson R (2008) Multi-scale Gaussian Mixtures for Cross-species Study ZiHua Yang 1: 1–14.

158. Pankla R, Buddhisa S, Berry M, Blankenship DM, Bancroft GJ, et al. (2009) Genomic transcriptional profiling identifies a candidate blood biomarker signature for the diagnosis of septicemic melioidosis. Genome

Biol 10: R127. doi:10.1186/gb-2009-10-11-r127.

159. How Kit A, Nielsen HM, Tost J (2012) DNA methylation based biomarkers: Practical considerations and applications. Biochimie 94: 2314–2337. doi:10.1016/j.biochi.2012.07.014.

160. Sivalingam SP, Sim SH, Jasper LCW, Wang D, Liu Y, et al. (2008) Pre- and post-exposure prophylaxis of experimental *Burkholderia pseudomallei* infection with doxycycline, amoxicillin/clavulanic acid and co-trimoxazole. J Antimicrob Chemother 61: 674–678. doi:10.1093/jac/dkm527.

161. Prucha M, Ruryk A, Boriss H, Möller E, Zazula R, et al. (2004) Expression profiling: toward an application in sepsis diagnostics. Shock 22: 29–33. doi:10.1097/01.shk.0000129199.30965.02.

162. Shanley TP, Cvijanovich N, Lin R, Allen GL, Thomas NJ, et al. (2007) Genome-level longitudinal expression of signaling pathways and gene networks in pediatric septic shock. Mol Med 13: 495–508. doi:10.2119/2007-00065.Shanley.

163. Pachot A, Lepape A, Vey S, Bienvenu J, Mougin B, et al. (2006) Systemic transcriptional analysis in survivor and non-survivor septic shock patients: A preliminary study. Immunol Lett 106: 63–71. doi:10.1016/j.imlet.2006.04.010.

164. Harris RA, Nagy-Szakal D, Pedersen N, Opekun A, Bronsky J, et al. (2012) Genome-wide peripheral blood leukocyte DNA methylation microarrays identified a single association with inflammatory bowel diseases. Inflamm Bowel Dis 18: 2334–2341. doi:10.1002/ibd.22956.

165. Kim M, Long TI, Arakawa K, Wang R, Yu MC, et al. (2010) DNA methylation as a biomarker for cardiovascular disease risk. PLoS One 5.

doi:10.1371/journal.pone.0009692.

166. Cassinotti E, Melson J, Liggett T, Melnikov A, Yi Q, et al. (2012) DNA methylation patterns in blood of patients with colorectal cancer and adenomatous colorectal polyps. Int J Cancer 131: 1153–1157. doi:10.1002/ijc.26484.

167. Florath I, Butterbach K, Heiss J, Bewerunge-Hudler M, Zhang Y, et al. (2016) Type 2 diabetes and leucocyte DNA methylation: an epigenome-wide association study in over 1,500 older adults. Diabetologia 59: 130–138. doi:10.1007/s00125-015-3773-7.

168. Pankla R, Buddhisa S, Berry M, Blankenship DM, Bancroft GJ, et al. (2009) Genomic transcriptional profiling identifies a candidate blood biomarker signature for the diagnosis of septicemic melioidosis. Genome Biol 10: R127. doi:10.1186/gb-2009-10-11-r127.

169. Chen YA, Lemire M, Choufani S, Butcher DT, Grafodatskaya D, et al. (2013) Discovery of cross-reactive probes and polymorphic CpGs in the Illumina Infinium HumanMethylation450 microarray. Epigenetics 8: 203–209. doi:10.4161/epi.23470.

170. Blair JD, Price EM (2012) Illuminating potential technical artifacts of DNA-methylation array probes. Am J Hum Genet 91: 760–762. doi:10.1016/j.ajhg.2012.05.028.

171. Jolliffe IT (2002) Principal Component Analysis, Second Edition. Encycl Stat Behav Sci 30: 487. doi:10.2307/1270093.

172. Horvath S (2013) DNA methylation age of human tissues and cell types. Genome biol 14: R115. doi:10.1186/gb-2013-14-10-r115.

173. Houseman EA, Accomando WP, Koestler DC, Christensen BC, Marsit CJ, et al. (2012) DNA methylation arrays as surrogate measures of cell

mixture distribution. BMC Bioinformatics 13: 86. doi:10.1186/1471-2105-13-86.

174. Phipson B, Maksimovic J, Oshlack A (2015) MissMethyl: An R package for analyzing data from Illumina's HumanMethylation450 platform. Bioinformatics 32: 286–288. doi:10.1093/bioinformatics/btv560.

175. Bock C (2012) Analysing and interpreting DNA methylation data. Nat Rev Genet 13: 705–719.

176. Brown MB (1975) A method for combining non-independent, one-sided tests of significance. Biometrics 31: pp. 987–992. doi:10.2307/2529826.

177. Fisher R (1925) Statistical methods for research workers. xv, 356 p. p. doi:10.1056/NEJMc061160.

178. Hess A, Iyer H (2007) Fisher's combined p-value for detecting differentially expressed genes using Affymetrix expression arrays. BMC Genomics 8: 96. doi:10.1186/1471-2164-8-96.

179. Bergsbaken T, Fink SL, Cookson BT (2009) Pyroptosis: host cell death and inflammation. Nat Rev Microbiol 7: 99–109. doi:10.1038/nrmicro2070.

180. Akira S, Takeda K (2004) Toll-like receptor signalling. Nat Rev Immunol 4: 499–511. doi:10.1038/nri1391.

181. Chen L, Flies DB (2013) Molecular mechanisms of T cell co-stimulation and co-inhibition. Nat Rev Immunol 13: 227–242. doi:10.1038/nri3405.

182. WONG KT, PUTHUCHEARY SD, VADIVELU J (1995) The histopathology of human melioidosis. Histopathology 26: 51–55. doi:10.1111/j.1365-2559.1995.tb00620.x.

183. Langfelder P, Horvath S (2008) WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics 9: 559. doi:10.1186/1471-2105-9-559.

184. Tobi EW, Lumey LH, Talens RP, Kremer D, Putter H, et al. (2009) DNA methylation differences after exposure to prenatal famine are common and timing- and sex-specific. Hum Mol Genet 18: 4046–4053. doi:10.1093/hmg/ddp353.

185. Kuehnen P, Mischke M, Wiegand S, Sers C, Horsthemke B, et al. (2012) An alu element-associated hypermethylation variant of the POMC gene is associated with childhood obesity. PLoS Genet 8. doi:10.1371/journal.pgen.1002543.

186. Maunakea AK, Chepelev I, Cui KR, Zhao KJ (2013) Intragenic DNA methylation modulates alternative splicing by recruiting MeCP2 to promote exon recognition. Cell Res 23: 1256–1269. doi:10.1038/cr.2013.110.

187. a P, L T, Am M, J L, Jl M, et al. (2015) Bacterial infection remodels the DNA methylation landscape of human dendritic cells. doi:10.1101/gr.192005.115.

188. Tan GMY, Looi CY, Fernandez KC, Vadivelu J, Loke MF, et al. (2015) Suppression of cell division-associated genes by *Helicobacter pylori* attenuates proliferation of RAW264.7 monocytic macrophage cells. Sci Rep 5: 11046. doi:10.1038/srep11046.

189. Pumirat P, Broek C Vander, Juntawieng N, Muangsombut V, Kiratisin P, et al. (2014) Analysis of the prevalence, secretion and function of a cell cycle-inhibiting factor in the melioidosis pathogen *Burkholderia pseudomallei*. PLoS One 9. doi:10.1371/journal.pone.0096298.

190. Storek KM, Monack DM (2015) Bacterial recognition pathways that lead to inflammasome activation. Immunol Rev 265: 112–129. doi:10.1111/imr.12289.

191. Bast A, Krause K, Schmidt IHE, Pudla M, Brakopp S, et al. (2014) Caspase-1-Dependent and -Independent Cell Death Pathways in *Burkholderia pseudomallei* Infection of Macrophages. PLoS Pathog 10. doi:10.1371/journal.ppat.1003986.

192. West TE, Myers ND, Chantratita N, Chierakul W, Limmathurotsakul D, et al. (2014) NLRC4 and TLR5 Each Contribute to Host Defense in Respiratory Melioidosis. PLoS Negl Trop Dis 8. doi:10.1371/journal.pntd.0003178.

193. Ceballos-Olvera I, Sahoo M, Miller MA, del Barrio L, Re F (2011) Inflammasome-dependent pyroptosis and IL-18 protect against burkholderia pseudomallei lung infection while IL-1β is deleterious. PLoS Pathog 7. doi:10.1371/journal.ppat.1002452.

194. Pierini R, Juruj C, Perret M, Jones CL, Mangeot P, et al. (2012) AIM2/ASC triggers caspase-8-dependent apoptosis in Francisella-infected caspase-1-deficient macrophages. Cell Death Differ 19: 1709–1721. doi:10.1038/cdd.2012.51.

195. Brinkmann V, Reichard U, Goosmann C, Fauler B, Uhlemann Y, et al. (2004) Neutrophil extracellular traps kill bacteria. Science (80-) 303: 1532–1535. doi:10.1126/science.1092385.

196. Riyapa D, Buddhisa S, Korbsrisate S, Cuccui J, Wren BW, et al. (2012) Neutrophil extracellular traps exhibit antibacterial activity against *Burkholderia pseudomallei* and are influenced by bacterial and host factors. Infect Immun 80: 3921–3929. doi:10.1128/IAI.00806-12.

197. de Jong HK, Koh G, Achouiti A, van der Meer AJ, Bulder I, et al. (2014) Neutrophil extracellular traps in the host defense against sepsis induced by *Burkholderia pseudomallei* (melioidosis). Intensive Care Med Exp 2:

21. doi:10.1186/s40635-014-0021-2.

198. Weehuizen TAF, Hommes TJ, Lankelma JM, de Jong HK, Roelofs JJTH, et al. (2016) Triggering Receptor Expressed on Myeloid Cells (TREM)-2 Impairs Host Defense in Experimental Melioidosis. PLoS Negl Trop Dis 10: e0004747. doi:10.1371/journal.pntd.0004747.

199. Sarovich DS, Price EP, Webb JR, Ward LM, Voutsinos MY, et al. (2014) Variable virulence factors in *Burkholderia pseudomallei* (Melioidosis) associated with human disease. PLoS One 9. doi:10.1371/journal.pone.0091682.

200. Essex-Lopresti AE, Boddey JA, Thomas R, Smith MP, Hartley MG, et al. (2005) A type IV pilin, pilA, contributes to adherence of Burkholderia pseudomallei and virulence in vivo. Infect Immun 73: 1260–1264. doi:10.1128/IAI.73.2.1260-1264.2005.

201. Stubben CJ, Duffield ML, Cooper I a, Ford DC, Gans JD, et al. (2009) Steps toward broad-spectrum therapeutics: discovering virulence-associated genes present in diverse human pathogens. BMC Genomics 10: 501. doi:10.1186/1471-2164-10-501.

202. Ooi WF, Ong C, Nandi T, Kreisberg JF, Chua HH, et al. (2013) The Condition-Dependent Transcriptional Landscape of Burkholderia pseudomallei. PLoS Genet 9. doi:10.1371/journal.pgen.1003795.

203. Smallbone K, Malys N, Messiha HL, Wishart JA, Simeonidis E (2011) Building a kinetic model of trehalose biosynthesis in *Saccharomyces cerevisiae.* Methods Enzymol 500: 355–370. doi:10.1016/B978-0-12-385118-5.00018-9.

204. van Heerden JH, Wortel MT, Bruggeman FJ, Heijnen JJ, Bollen YJM, et al. (2014) Lost in Transition: Startup of Glycolysis Yields Subpopulations

of Nongrowing Cells. Science (80-) 343: 1245114–1245114. doi:10.1126/science.1245114.

205. Planqué R, Bruggeman FJ, Teusink B, Hulshof J (2014) Understanding bistability in yeast glycolysis using general properties of metabolic pathways. Math Biosci 255: 33–42. doi:10.1016/j.mbs.2014.06.006.

206. Thevelein JM, Hohmann S (1995) Trehalose synthase: guard to the gate of glycolysis in yeast? Trends Biochem Sci 20: 3–10. doi:10.1016/S0968-0004(00)88938-0.

207. Blázquez MA, Lagunas R, Gancedo C, Gancedo JM (1993) Trehalose-6-phosphate, a new regulator of yeast glycolysis that inhibits hexokinases. FEBS Lett 329: 51–54. doi:10.1016/0014-5793(93)80191-V.

208. Teusink B, Walsh MC, Van Dam K, Westerhoff H V. (1998) The danger of metabolic pathways with turbo design. Trends Biochem Sci 23: 162–169. doi:10.1016/S0968-0004(98)01205-5.

209. van Heerden JH, Wortel MT, Bruggeman FJ, Heijnen JJ, Bollen YJM, et al. (2014) Lost in transition: start-up of glycolysis yields subpopulations of nongrowing cells. Science 343: 1245114. doi:10.1126/science.1245114.

210. Harrison MAA (2009) Encyclopedia of Microbiology. 23-33 p. doi:10.1016/B978-012373944-5.00117-6.

211. O'Leary NA, Wright MW, Brister JR, Ciufo S, Haddad D, et al. (2016) Reference sequence (RefSeq) database at NCBI: Current status, taxonomic expansion, and functional annotation. Nucleic Acids Res 44: D733–D745. doi:10.1093/nar/gkv1189.

212. Winsor GL, Khaira B, Van Rossum T, Lo R, Whiteside MD, et al. (2008) The Burkholderia Genome Database: Facilitating flexible queries and comparative analyses. Bioinformatics 24: 2803–2804.

doi:10.1093/bioinformatics/btn524.

213. Cherry JM, Hong EL, Amundsen C, Balakrishnan R, Binkley G, et al. (2012) Saccharomyces Genome Database: The genomics resource of budding yeast. Nucleic Acids Res 40. doi:10.1093/nar/gkr1029.

214. Zhou J, Rudd KE (2013) EcoGene 3.0. Nucleic Acids Res 41. doi:10.1093/nar/gks1235.

215. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, et al. (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol Syst Biol 7: 539. doi:10.1038/msb.2011.75.

216. Rost B (1999) Twilight zone of protein sequence alignments. Protein Eng 12: 85–94. doi:10.1093/protein/12.2.85.

217. Marchler-Bauer A, Derbyshire MK, Gonzales NR, Lu S, Chitsaz F, et al. (2015) CDD: NCBI's conserved domain database. Nucleic Acids Res 43: D222–D226. doi:10.1093/nar/gku1221.

218. Pearson WR (2013) An introduction to sequence similarity ("homology") searching. Curr Protoc Bioinforma. doi:10.1002/0471250953.bi0301s42.

219. Michaelis L, Menten ML (1913) Die Kinetik der Invertinwirkung. Biochem Z 49: 333–369. doi:10.1021/bi201284u.

220. Klipp E, Liebermeister W, Wierling C, Kowald A, Lehrach H, et al. (2009) Systems Biology: A Textbook. 569 p. doi:10.3797/scipharm.

221. Wattam AR, Abraham D, Dalay O, Disz TL, Driscoll T, et al. (2014) PATRIC, the bacterial bioinformatics database and analysis resource. Nucleic Acids Res 42. doi:10.1093/nar/gkt1099.

222. Rodrigues F, Sarkar-Tyson M, Harding S V., Siew HS, Hui HC, et al. (2006) Global map of growth-regulated gene expression in *Burkholderia pseudomallei*, the causative agent of melioidosis. J Bacteriol 188: 8178–

8188. doi:10.1128/JB.01006-06.

223. Devoid S, Overbeek R, DeJongh M, Vonstein V, Best AA, et al. (2013) Automated genome annotation and metabolic model reconstruction in the SEED and model SEED. Methods Mol Biol 985: 17–45. doi:10.1007/978-1-62703-299-5-2.

224. Schellenberger J, Que R, Fleming RMT, Thiele I, Orth JD, et al. (2011) Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox v2.0. Nat Protoc 6: 1290–1307. doi:10.1038/nprot.2011.308.

225. Barraza A, Sánchez F (2013) Trehalases: a neglected carbon metabolism regulator? Plant Signal Behav 8: e24778. doi:10.4161/psb.24778.

226. Hohmann S, Bell W, Neves MJ, Valckx D, Thevelein JM (1996) Evidence for trehalose-6-phosphate-dependent and -independent mechanisms in the control of sugar influx into yeast glycolysis. Mol Microbiol 20: 981–991. doi:10.1111/j.1365-2958.1996.tb02539.x.

227. Engel SR, Dietrich FS, Fisk DG, Binkley G, Balakrishnan R, et al. (2014) The reference genome sequence of *Saccharomyces cerevisiae*: then and now. G3 (Bethesda) 4: 389–398. doi:10.1534/g3.113.008995.

228. Postma PW, Lengeler JW, Jacobson GR (1993) Phosphoenolpyruvate: Carbohydrate phosphotransferase systems of bacteria. Microbiol Rev 57: 543–594.

229. Wilson JE (1995) Hexokinases. Rev Physiol Biochem Pharmacol 126: 65–198.

230. Meyer D, Schneider-Fresenius C, Horlacher R, Peist R, Boos W (1997) Molecular characterization of glucokinase from Escherichia coli K-12. J Bacteriol 179: 1298–1306.

231. CABIB E, LELOIR LF (1958) The biosynthesis of trehalose phosphate. J Biol Chem 231: 259–275. doi:10.1002/1097-0142(195707/08)10:4<711::AID-CNCR2820100412>3.0.CO;2-7.

232. Pritchard L, Kell DB (2002) Schemes of flux control in a model of *Saccharomyces cerevisiae* glycolysis. Eur J Biochem 269: 3894–3904. doi:10.1046/j.1432-1033.2002.03055.x.

233. Vandercammen a, François J, Hers HG (1989) Characterization of trehalose-6-phosphate synthase and trehalose-6-phosphate phosphatase of *Saccharomyces cerevisiae*. Eur J Biochem 182: 613–620. doi:10.1111/j.1432-1033.1989.tb14870.x.

234. Wright BE, Killick KA, Tai A (1977) Fourth expansion and glucose perturbation of the Dictyostelium kinetic model. Eur J Biochem 74: 217–225.

235. Ernandes JR, De Meirsman C, Rolland F, Winderickx J, De Winde J, et al. (1998) During the initiation of fermentation overexpression of hexokinase PII in yeast transiently causes a similar deregulation of glycolysis as deletion of Tps1. Yeast 14: 255–269. doi:10.1002/(SICI)1097-0061(199802)14:3<255::AID-YEA228>3.0.CO;2-N.

236. Kresnowati MTAP, Van Winden WA, Almering MJH, ten Pierick A, Ras C, et al. (2006) When transcriptome meets metabolome: fast cellular responses of yeast to sudden relief of glucose limitation. Mol Syst Biol 2: 49. doi:10.1038/msb4100083.

237. Schuetz R, Kuepfer L, Sauer U (2007) Systematic evaluation of objective functions for predicting intracellular fluxes in Escherichia coli. Mol Syst Biol 3: 119. doi:10.1038/msb4100162.

238. Moule MG, Hemsley CM, Seet Q, Guerra-Assuno JA, Lim J, et al. (2014) Genome-wide saturation mutagenesis of *Burkholderia pseudomallei* K96243 predicts essential genes and novel targets for antimicrobial development. MBio 5. doi:10.1128/mBio.00926-13.

239. Nogales J, Palsson BØ, Thiele I (2008) A genome-scale metabolic reconstruction of Pseudomonas putida KT2440: iJN746 as a cell factory. BMC Syst Biol 2: 79. doi:10.1186/1752-0509-2-79.

240. Thiele I, Vo TD, Price ND, Palsson BØ (2005) Expanded metabolic reconstruction of *Helicobacter pylori* (iIT341 GSM/GPR): An in silico genome-scale characterization of single- and double-deletion mutants. J Bacteriol 187: 5818–5830. doi:10.1128/JB.187.16.5818-5830.2005.

241. Erdrich P, Steuer R, Klamt S (2015) An algorithm for the reduction of genome-scale metabolic network models to meaningful core models. BMC Syst Biol 9: 1–12. doi:10.1186/s12918-015-0191-x.

242. Schuster S, Fell DA, Dandekar T (2000) A general definition of metabolic pathways useful for systematic organization and analysis of complex metabolic networks. Nat Biotechnol 18: 326–332. doi:10.1038/73786.

243. Chaussabel D, Pascual V, Banchereau J (2010) Assessing the human immune system through blood transcriptomics. BMC Biol 8: 84. doi:10.1186/1741-7007-8-84.

244. Chaussabel D, Quinn C, Shen J, Patel P, Glaser C, et al. (2008) A modular analysis framework for blood genomics studies: application to systemic lupus erythematosus. Immunity 29: 150–164. doi:10.1016/j.immuni.2008.05.012.

245. Liu J, Chen J, Ehrlich S, Walton E, White T, et al. (2014) Methylation patterns in whole blood correlate with symptoms in Schizophrenia

patients. Schizophr Bull 40: 769–776. doi:10.1093/schbul/sbt080.

246. Teschendorff AE, Yang Z, Wong A, Pipinikas CP, Jiao Y, et al. (2015) Correlation of Smoking-Associated DNA Methylation Changes in Buccal Cells With DNA Methylation Changes in Epithelial Cancer. JAMA Oncol 1: 476–485. doi:10.1001/jamaoncol.2015.1053.

247. Bernstein BE, Stamatoyannopoulos JA, Costello JF, Ren B, Milosavljevic A, et al. (2010) The NIH Roadmap Epigenomics Mapping Consortium. Nat Biotechnol 28: 1045–1048. doi:10.1038/nbt1010-1045.

248. Chandra G, Chater KF, Bornemann S (2011) Unexpected and widespread connections between bacterial glycogen and trehalose metabolism. Microbiology 157: 1565–1572. doi:10.1099/mic.0.044263-0.

APPENDIX A

List of differentially methylated regions.

Chromosome	Start	End	Number of CpGs in the region	Gene	P-value
16	6623995	6624293	4	A2BP1	5.95E-03
10	27150052	27150112	3	ABI1	3.30E-03
11	34378797	34378975	6	ABTB2	4.67E-03
16	89219415	89219547	3	ACSF3	8.89E-03
10	1446913	1447183	3	ADARB2	5.86E-03
16	4013337	4013659	3	ADCY9	2.29E-03
1	112058236	112058397	3	ADORA3	7.81E-03
5	314553	314642	4	AHRR;PDCD6	3.77E-03
11	46612744	46613132	3	AMBRA1	7.87E-03
3	49757306	49757476	4	AMIGO3;RNF123	6.36E-03
5	112042851	112042966	4	APC	9.12E-03
13	107220428	107220746	7	ARGLU1	7.46E-03
8	17942444	17942569	4	ASAH1	7.47E-03
2	85980500	85980609	3	ATOH8	6.12E-03
6	33244799	33245181	6	B3GALT4	8.48E-03
7	33168625	33168904	3	BBS9	7.58E-03
20	49411412	49411594	3	BCAS4	6.42E-03
11	27722971	27723290	8	BDNF	3.35E-03
11	27723128	27723409	8	BDNF	4.31E-03
11	27723075	27723385	8	BDNF	4.66E-03
11	27723190	27723409	7	BDNF	7.58E-03
2	28113413	28113747	5	BRE;RBKS;LOC100302650	7.38E-03
14	60558485	60558674	4	C14orf135	4.96E-03
21	46131601	46131737	3	C21orf29	5.94E-03
22	32555255	32555310	4	C22orf42	9.57E-03
8	143808262	143808440	3	C8orf55	4.83E-04
16	1204876	1205018	3	CACNA1H	2.60E-03
5	179107813	179108129	3	CBY3	1.95E-03
12	111284527	111284768	5	CCDC63	9.81E-03
16	3086292	3086538	3	CCDC64B	3.49E-03
11	60609242	60609393	4	CCDC86	9.26E-03
20	3776748	3776985	3	CDC25B	8.42E-03
1	111770111	111770291	3	CHI3L2	9.42E-03
21	37915044	37915281	4	CLDN14	5.22E-03
21	37915091	37915391	4	CLDN14	8.71E-03
21	37915044	37915091	3	CLDN14	9.67E-03
16	28503005	28503160	3	CLN3	3.59E-04
16	28503005	28503311	4	CLN3	2.61E-03
4	17516493	17516739	4	CLRN2	6.96E-04
4	17516493	17516661	3	CLRN2	1.11E-03
4	17516493	17516819	6	CLRN2	2.24E-03
18	500979	501234	4	COLEC12	1.54E-03
18	500817	501137	4	COLEC12	3.55E-03
4	84205972	84206199	5	COQ2	7.91E-03
1	207627372	207627541	3	CR2	7.24E-03
6	31635781	31635946	3	CSNK2B	1.56E-03
6	31635781	31636140	4	CSNK2B	6.51E-03
20	23433349	23433608	3	CST11	4.64E-03
11	118763859	118764076	6	CXCR5	7.44E-03
19	6482011	6482169	3	DENND1C	3.86E-03
12	16064140	16064282	4	DERA	6.14E-03
12	125473657	125474021	6	DHX37	7.48E-03
12	125473597	125473719	7	DHX37	9.28E-03
10	444914	445130	3	DIP2C	4.55E-03
11	84431850	84431970	3	DLG2	5.66E-03
2	172965046	172965280	3	DLX2	4.88E-03
5	13810143	13810279	3	DNAH5	5.79E-03

List of differentially methylated regions (cont.).

Chromosome	Start	End	Number of CpGs in the region	Gene	P-value
17	11501999	11502114	3	DNAH9	9.74E-03
11	117665395	117665622	3	DSCAML1	5.42E-03
6	31855269	31855457	6	EHMT2	1.86E-07
6	31855269	31855498	7	EHMT2	2.38E-07
6	31855269	31855598	8	EHMT2	1.97E-06
6	31855381	31855695	8	EHMT2	6.00E-06
6	31855422	31855709	9	EHMT2	1.93E-05
17	39844656	39844988	5	EIF1	5.23E-03
17	39844834	39845092	7	EIF1	5.84E-03
14	100259801	100259860	3	EML1	2.72E-03
12	7032793	7033146	3	ENO2;ATN1	9.18E-03
17	48609923	48610110	4	EPN3	3.18E-03
17	48609747	48610110	5	EPN3	7.11E-03
7	14030351	14030482	3	ETV1	1.49E-03
7	14030255	14030482	4	ETV1	3.84E-03
1	93258288	93258319	3	EVI5	8.58E-04
1	11160013	11160205	8	EXOSC10	6.54E-03
1	118166117	118166402	3	FAM46C	2.57E-03
15	59062977	59063176	3	FAM63B	3.23E-03
5	126409372	126409573	4	FLJ44606	3.23E-03
5	126409308	126409573	6	FLJ44606	7.12E-03
1	100231977	100232077	3	FRRS1	1.14E-03
7	30634387	30634502	4	GARS	8.99E-03
1	94374966	94375328	6	GCLM	6.72E-03
5	179779115	179779237	3	GFPT2	3.94E-03
14	54955792	54955960	5	GMFB	7.10E-05
7	50660435	50660646	3	GRB10	6.19E-03
4	7069733	7069943	7	GRPEL1	1.88E-03
4	7069733	7069911	5	GRPEL1	3.94E-03
4	7069733	7070051	8	GRPEL1	4.18E-03
4	7069857	7070051	5	GRPEL1	6.11E-03
4	7069785	7070051	7	GRPEL1	6.48E-03
9	124030411	124030508	3	GSN	7.03E-03
11	76433173	76433300	3	GUCY2E	3.43E-03
6	27783054	27783331	3	HIST1H2AJ;HIST1H2BM	7.46E-03
6	32975590	32975885	7	HLA-DOA	8.36E-03
3	185541074	185541217	3	IGF2BP2	1.45E-03
3	185541074	185541276	4	IGF2BP2	3.73E-03
11	133816022	133816097	3	IGSF9B	1.87E-03
11	133816022	133816242	4	IGSF9B	3.94E-03
3	10206254	10206537	6	IRAK2	3.87E-03
3	10206254	10206413	5	IRAK2	6.67E-03
3	10206344	10206731	7	IRAK2	8.60E-03
7	139877050	139877104	3	JHDM1D;LOC100134229	6.25E-03
1	215178658	215178907	3	KCNK2	2.52E-03
11	2466598	2466909	5	KCNQ1	4.59E-03
8	99439685	99439948	3	KCNS2	1.07E-04
19	5147855	5148130	3	KDM4B	9.32E-03
6	30653407	30653659	5	KIAA1949	9.63E-03
1	245317021	245317174	3	KIF26B	2.37E-03
1	245317789	245317980	4	KIF26B	7.22E-03
1	245316885	245317174	4	KIF26B	8.37E-03
19	55013915	55014066	4	LAIR2	7.56E-03
3	49170599	49170794	4	LAMB2	2.33E-03
3	49170496	49170794	5	LAMB2	2.47E-03
3	49170496	49170668	3	LAMB2	5.92E-03
19	35615444	35615638	3	LGI4	7.23E-03

List of differentially methylated regions (cont.).

Chromosome	Start	End	Number of CpGs in the region	Gene	P-value
1	180204974	180205253	3	LHX4	8.35E-03
17	41467024	41467212	3	LOC100130581	1.45E-03
17	41466953	41467212	4	LOC100130581	9.71E-03
3	10806175	10806370	4	LOC285370	3.87E-03
3	10806114	10806370	6	LOC285370	5.15E-03
3	10806022	10806370	8	LOC285370	6.40E-03
3	10806047	10806370	7	LOC285370	6.93E-03
1	99470618	99470801	3	LPPR5	4.26E-03
4	151935870	151935948	3	LRBA	9.57E-03
2	238608549	238608716	3	LRRFIP1	9.72E-03
19	34662830	34663190	5	LSM14A	7.31E-03
7	1986245	1986334	3	MAD1L1	2.25E-05
7	1986245	1986533	4	MAD1L1	6.16E-05
7	1986334	1986640	3	MAD1L1	7.64E-04
7	2057703	2058031	3	MAD1L1	9.38E-03
4	156298264	156298427	3	MAP9	6.83E-03
18	58039932	58040021	3	MC4R	6.29E-03
18	58039932	58040213	4	MC4R	8.38E-03
19	42873746	42873882	3	MEGF8	7.59E-03
7	7607013	7607056	3	MIOS	5.71E-03
11	121971092	121971247	4	MIR125B1;LOC399959	2.08E-03
14	101489985	101490145	6	MIR299;MIR380	3.45E-03
14	101490118	101490253	5	MIR299;MIR380	4.68E-03
14	101489985	101490186	7	MIR299;MIR380	8.70E-03
11	58980377	58980665	3	MPEG1	4.35E-04
11	68782049	68782202	3	MRGPRF	6.93E-04
11	68781976	68782106	3	MRGPRF	1.07E-03
11	68781976	68782202	4	MRGPRF	1.57E-03
1	155162871	155163205	3	MUC1	4.19E-03
16	3507546	3507875	5	NAT15	3.65E-03
16	3507460	3507735	6	NAT15	5.20E-03
16	3507460	3507582	5	NAT15	7.07E-03
11	67373977	67374213	3	NDUFV1	7.51E-03
14	24836084	24836219	5	NFATC4	1.62E-03
14	32030205	32030546	4	NUBPL	4.14E-03
7	44530556	44530730	3	NUDCD3	8.85E-03
16	56764314	56764501	3	NUP93	8.84E-03
1	228403357	228403515	3	OBSCN	8.82E-04
1	228403247	228403515	4	OBSCN	3.15E-03
6	29394438	29394654	3	OR11A1	3.75E-03
17	72931569	72931792	7	OTOP3	5.05E-03
17	72931569	72931857	10	OTOP3	5.45E-03
17	72931569	72931903	11	OTOP3	5.75E-03
11	65983765	65983930	3	PACS1	2.24E-03
11	65983930	65984211	3	PACS1	3.99E-03
11	65983640	65983930	4	PACS1	7.76E-03
11	65983765	65984109	4	PACS1	8.93E-03
10	3168669	3168776	3	PFKP	3.00E-03
11	74109729	74109971	3	PGM2L1	7.97E-03
9	139743610	139743802	3	PHPT1	7.74E-03
22	38470264	38470483	3	PICK1	6.38E-03
5	67586170	67586433	3	PIK3R1	7.41E-03
12	57984650	57984821	5	PIP4K2C	7.22E-03
10	3197650	3197757	3	PITRM1	5.41E-03
4	111551642	111551987	3	PITX2	4.41E-03
15	42302342	42302500	3	PLA2G4E	4.27E-03
1	2428080	2428333	3	PLCH2	5.94E-03

List of differentially methylated regions (cont.).

Chromosome	Start	End	Number of CpGs in the region	Gene	P-value
1	6546646	6546839	3	PLEKHG5	4.61E-03
1	6537569	6537785	3	PLEKHG5	5.22E-03
14	20937107	20937394	5	PNP	6.77E-03
3	127347876	127348155	5	PODXL2	4.26E-04
3	127347536	127347884	5	PODXL2	1.22E-03
3	127347685	127347978	5	PODXL2	1.79E-03
5	6755150	6755439	5	POLS	4.77E-03
5	6755235	6755439	4	POLS	5.14E-03
5	6755265	6755439	3	POLS	5.50E-03
5	6755026	6755265	4	POLS	6.02E-03
19	291239	291315	5	PPAP2C	4.07E-03
11	70211301	70211531	3	PPFIA1	1.86E-03
14	50065185	50065333	6	PPIL5	6.95E-03
12	63328524	63328717	4	PPM1H	9.08E-03
1	3144536	3144710	3	PRDM16	4.26E-03
10	72358510	72358696	3	PRF1	4.42E-03
12	120105695	120105824	3	PRKAB1	8.09E-03
5	177423311	177423379	3	PROP1	1.30E-03
7	29603121	29603253	3	PRR15	7.76E-03
18	12884517	12884567	3	PTPN2	7.79E-03
7	157690609	157690769	4	PTPRN2	8.36E-04
7	157412694	157413057	4	PTPRN2	5.31E-03
7	157340390	157340554	3	PTPRN2	7.89E-03
7	4850105	4850345	3	RADIL	6.71E-03
2	178976981	178977094	4	RBM45	9.30E-04
2	178976981	178977215	5	RBM45	5.90E-03
1	17751614	17751920	3	RCC2	8.73E-03
1	17751920	17752023	3	RCC2	9.56E-03
17	33391738	33391960	3	RFFL	1.10E-03
6	72924771	72925046	3	RIMS1	9.99E-03
15	59279406	59279703	3	RNF111	4.15E-03
17	5322937	5323193	7	RPAIN;NUP88	1.15E-03
17	5322836	5323110	7	RPAIN;NUP88	7.87E-03
6	35435693	35436047	5	RPL10A	9.42E-04
6	35435693	35435861	3	RPL10A	9.36E-03
8	99058424	99058672	3	RPL30	7.70E-03
20	35807784	35807911	3	RPN2;C20orf132	3.94E-03
10	92631428	92631716	8	RPP30	6.70E-03
17	78682785	78683082	3	RPTOR	1.42E-03
17	27312855	27313157	4	SEZ6	2.99E-03
7	101936230	101936430	4	SH2B2	3.68E-03
7	101936230	101936527	5	SH2B2	3.88E-03
7	101936362	101936527	3	SH2B2	5.86E-03
2	110259084	110259223	3	SH3RF3	7.97E-03
11	70564116	70564414	3	SHANK2	1.38E-03
11	70516997	70517057	3	SHANK2	9.53E-03
7	155597344	155597490	3	SHH	3.22E-03
5	1212434	1212554	3	SLC6A19	6.33E-03
3	181429282	181429654	6	SOX2OT;SOX2	9.34E-03
2	231090531	231090745	3	SP140	7.57E-03
4	184938291	184938470	4	STOX2	1.85E-03
17	61678300	61678664	3	TACO1	9.22E-03
2	162272107	162272299	3	TBR1	4.56E-03
10	102890038	102890296	3	TLX1;TLX1NB	8.51E-03
17	8079766	8079962	7	TMEM107	8.18E-03
10	104221031	104221224	3	TMEM180	2.83E-03
20	44451667	44451973	4	TNNC2	9.28E-03

List of differentially methylated regions (cont.).

Chromosome	Start	End	Number of CpGs in the region	Gene	P-value
2	218770208	218770270	3	TNS1	9.03E-03
14	58862146	58862417	3	TOMM20L	1.69E-03
8	144403126	144403439	4	TOP1MT	3.72E-03
11	128813442	128813583	4	TP53AIP1	7.11E-03
3	189348936	189349323	6	TP63	1.16E-03
3	189348936	189349129	4	TP63	2.05E-03
3	189349129	189349323	3	TP63	7.19E-03
6	28889876	28890069	4	TRIM27	4.39E-04
6	28889682	28890069	5	TRIM27	4.66E-04
7	99517136	99517295	8	TRIM4	5.23E-03
7	99517277	99517509	8	TRIM4	6.94E-03
7	99517192	99517509	10	TRIM4	7.10E-03
7	99517136	99517460	11	TRIM4	7.98E-03
7	99517136	99517378	10	TRIM4	8.18E-03
5	180650178	180650516	5	TRIM41	6.31E-03
5	14508132	14508414	3	TRIO	5.61E-03
5	176073997	176074033	3	TSPAN17	6.21E-03
3	50365842	50366021	4	TUSC2	8.04E-03
1	154192352	154192647	4	UBAP2L;C1orf43	7.25E-03
6	33231322	33231631	3	VPS52	1.17E-03
6	33231505	33231813	4	VPS52	1.99E-03
11	75898132	75898203	3	WNT11	1.02E-03
2	113012287	113012520	3	ZC3H8	5.62E-03
4	2372761	2372960	3	ZFYVE28	6.80E-03
1	244218327	244218537	3	ZNF238	1.53E-03
19	38085706	38085810	6	ZNF540;ZNF571	4.85E-03
1	91486951	91487159	3	ZNF644	6.43E-03
20	47874072	47874155	3	ZNFX1	3.49E-03
19	58630089	58630429	3	ZSCAN18	2.18E-03
1	33938117	33938307	6	ZSCAN20	7.91E-03

APPENDIX B

SCIENTIFIC REPORTS

OPEN

Received: 19 April 2016
Accepted: 07 July 2016
Published: 03 August 2016

Mapping epigenetic changes to the host cell genome induced by *Burkholderia pseudomallei* reveals pathogen-specific and pathogen-generic signatures of infection

Deniz Cizmeci[1,*], Emma L. Dempster[2,*], Olivia L. Champion[3], Sariqa Wagley[3], Ozgur E. Akman[1], Joann L. Prior[3], Orkun S. Soyer[4], Jonathan Mill[2,5] & Richard W. Titball[3]

The potential for epigenetic changes in host cells following microbial infection has been widely suggested, but few examples have been reported. We assessed genome-wide patterns of DNA methylation in human macrophage-like U937 cells following infection with *Burkholderia pseudomallei*, an intracellular bacterial pathogen and the causative agent of human melioidosis. Our analyses revealed significant changes in host cell DNA methylation, at multiple CpG sites in the host cell genome, following infection. Infection induced differentially methylated probes (iDMPs) showing the greatest changes in DNA methylation were found to be in the vicinity of genes involved in inflammatory responses, intracellular signalling, apoptosis and pathogen-induced signalling. A comparison of our data with reported methylome changes in cells infected with *M. tuberculosis* revealed commonality of differentially methylated genes, including genes involved in T cell responses (*BCL11B*, *FOXO1*, *KIF13B*, *PAWR*, *SOX4*, *SYK*), actin cytoskeleton organisation (*ACTR3*, *CDC42BPA*, *DTNBP1*, *FERMT2*, *PRKCZ*, *RAC1*), and cytokine production (*FOXP1*, *IRF8*, *MR1*). Overall our findings show that pathogenic-specific and pathogen-common changes in the methylome occur following infection.

Melioidosis is an infectious disease caused by the intracellular bacterial pathogen *Burkholderia pseudomallei*. A recent study has found that the incidence of this disease is likely to have been significantly underestimated[1]. *B. pseudomallei* adopts various strategies to survive and replicate within host cells[2]. Avoiding or resisting the antimicrobial activities in the phagosome of host cells[2,3] allows escape into the cytosol, the induction of actin polymerization and cell to cell spreading[4]. The transcriptional response of humans or mice to *B. pseudomallei* infection reveals changes in the expression of multiple genes, including loci associated with the inflammatory and innate immune responses[5–7]. These changes might be driven actively by bacteria, or they may be the consequence of altered responses of host cells to the infection.

Recently, a role for epigenetic regulation of host cell function during bacterial infection has been suggested[8,9]. DNA methylation, occurring exclusively at cytosine residues in mammals, is one of the best understood epigenetic mechanisms, playing a key role in transcriptional regulation during development and being increasingly implicated in a range of non-infectious diseases[10]. Epigenetic processes can be dynamic: they are influenced upon exposure to a range of external environmental factors and stochastic events in the cell[11].

Few studies have investigated epigenetic changes in response to infection but there are reports of altered patterns on methylation in cultured cells infected with *Helicobacter pylori*[12,13], *Mycobacterium tuberculosis*[14] or *Leishmania donovani*[15] and evidence that in mice the gut flora influences methylation of the *IL-4* gene in intestinal

[1]College of Engineering, Mathematics and Physical Sciences, University of Exeter, Exeter, United Kingdom. [2]University of Exeter Medical School, Exeter University, Exeter, United Kingdom. [3]College of Life and Environmental Sciences, University of Exeter, Exeter, United Kingdom. [4]School of Life Sciences, University of Warwick, United Kingdom. [5]Institute of Psychiatry, Psychology & Neuroscience, King's College London, United Kingdom. *These authors contributed equally to this work. Correspondence and requests for materials should be addressed to O.S.S. (email: o.soyer@warwick.ac.uk) or J.M. (email: j.mill@exeter.ac.uk) or R.W.T. (email: r.w.titball@exeter.ac.uk)

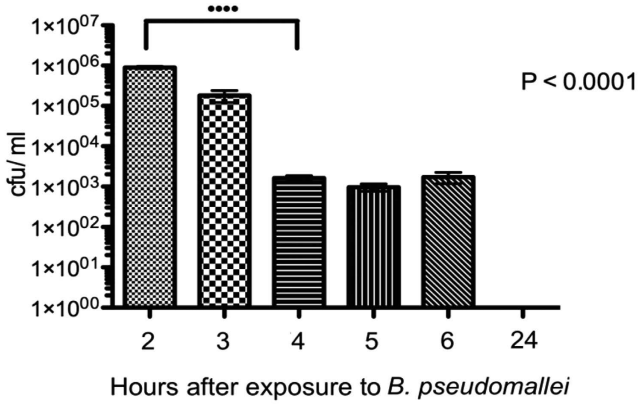


Figure 1. Bacterial loads at different time points in U937 cells infected with *B. pseudomallei* (MOI = 10). Bacterial load was measured as colony forming units (CFU).

epithelial cells resulting in the down-regulation of *TLR4* expression[16]. Some of these studies have focussed on selected genes[12,16], others have profiled genome wide changes in the methylome[13–15]. These studies have revealed linkage between methylation events and the level of expression of associated gene(s).

In this study we determined genome-wide changes in DNA methylation in human macrophage-like U937 cells following infection by *B. pseudomallei*. We identify widespread changes in host cell DNA methylation following infection, and show that these are enriched in the vicinity of loci involved in inflammatory responses, intracellular signalling, apoptosis and pathogen-induced signalling. By comparing our data with the previously reported data we have identified genes, which show differential patterns of methylation in cells infected with different pathogens.

Results

Infection Model. U937 cells can mature and differentiate to adopt the characteristics of mature human macrophages[17]. We established the pattern of infection of U937 cells, which had been activated with interferon gamma (IFN-γ), by *B. pseudomallei* K96243 expressing red fluorescent protein (RFP). At 2 hours (T2) post infection, 74% of the human cells were infected (Supplementary Fig. 1) and the number of intracellular bacteria was 10^6 CFU/ml (Fig. 1). After 4 (T4) or 6 hours (T6) post infection the number of intracellular bacteria had declined to 10^3 CFU/ml. No intracellular bacteria were recovered at 24 hours post infection (T24). We and others[18] have found that the activation of macrophages with IFN-γ dramatically enhances the ability of the cells to control infection with *B. pseudomallei* and this explains the progressive reduction in the number of intracellular bacteria.

Multiple loci in U937 cells show reproducible changes in DNA methylation after B. pseudomallei infection. We infected U937 cells with *B. pseudomallei* and mapped changes in host cell DNA methylation at T2 and T4. DNA from the infected or uninfected (control) U937 cells was collected from two technical replicates and two experimental replicates at each time. We subsequently performed a second experiment following the same protocol but with additional sampling times included to provide samples at 1 hour (T1), 2 hours (T2), 3 hours (T3) and 4 hours (T4) post infection. The experimental design is shown in Supplementary Fig. 2.

DNA methylation was quantified using an Illumina 450K HumanMethylation array, with pre-processing, normalization and stringent quality control undertaken. In our first experiment we identified differentially-methylated positions (DMPs) between infected and uninfected cells at T2 or T4, allowing us to identify infection-induced DMPs (iDMPs).

Reasoning that larger differences are potentially more biologically meaningful, our primary focus was on iDMPs characterized by >10% DNA methylation difference between groups[19]. We identified 10,279 iDMPs (54% hypo-methylated, 46% hyper-methylated) at T2 and 4850 iDMPs (57% hypo-methylated, 43% hyper-methylated) at T4, with 642 iDMPs differentially methylated at both T2 and T4 (Fig. 2). We next performed a second experiment to identify consistent changes and obtained a stringently-filtered dataset of 388 conserved iDMPs conserved between the two experiments and which were used for subsequent analysis. These DNA methylation changes were highly correlated across both experiments (T2: $r = 0.83$, $p < 0.0001$, Supplementary Fig. 3a; T4: $r = 0.73$, $p < 0.0001$, Supplementary Fig. 3b). 264 iDMPs were identical at T2, 141 iDMPs were conserved at T4 and 17 iDMPs conserved at both T2 and T4 (Supplementary Table 1).

The conserved iDMPs were not uniformly distributed across the genome (Supplementary Table 2). We found a highly-significant over- representation of infection-induced DNA methylation at probes located in the first exons of genes (compared to the frequency of positions in all probes on the Illumina 450K array analysed, relative enrichment (95% CI) = 3.34, $P = 1.11 \times 10^{-22}$).

In addition to quantifying DNA methylation at T2 and T4 in experiment 1 and 2, we profiled samples collected at T1 and T3 in experiment 2 (Supplementary Fig. 2). Distinct patterns of DNA methylation across the four time points were observed (Table 1); 21 probes (5.41%) showed consistent changes in DNA methylation across all time points in both experiments (for example cg17676428 (Fig. 3a)), 23 (5.93%) showed large changes at the early time-points that diminished at the later time-points (for example cg15470658 (Fig. 3b)), 55 (14.18%) displayed a

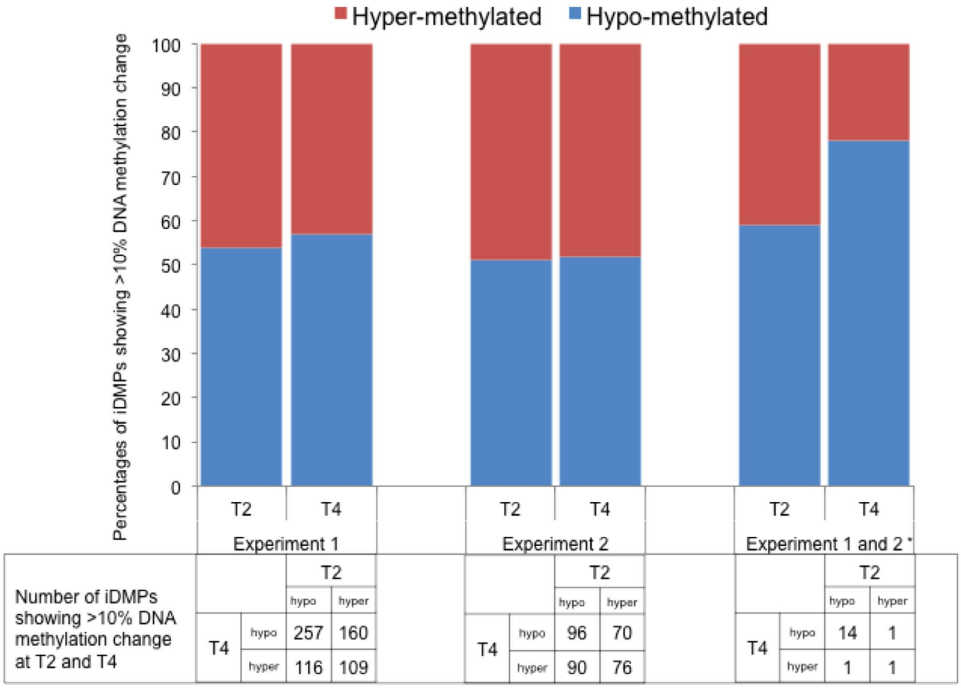


Figure 2. Quantification of Infection-induced iDMPs. *iDMPs conserved between the two experiments.

Category	Number of probes	Hyper-methylated	Hypo-methylated
A: constant response	21	3	18
B: early response	23	10	13
C: late response	55	11	44
D: transient response	99	42	57
E: oscillatory response	190	NA	NA

Table 1. Number of iDMPs assigned to categories based on differential methylation patterns. Signatures with differential methylation at all time points (1 hour (T1), 2 hours (T2), 3 hours (T3) and 4 hours (T4) post infection) were classified as, constant hyper- or constant hypo-methylated probes. Signatures with differential methylation at early time points (T1, T2, T3) and no change at the T4 time point were classified as early response probes. Signatures with no response at the T1 time point and with differential methylation at later time points were classified as late response probes. Signatures with no response at T1 and T4 and differential methylation at T2 or T3 were classified as transient responses. The remainder of the patterns were classified as oscillatory probes. NA = not applicable.

lag in response, with DNA methylation changes only occurring at the later time points after infection (for example cg14113958 (Fig. 3c)), 99 (25.52%) showed a transient (for example cg14173033 (Fig. 3d)), and 190 (48.97%) an oscillatory response. The losses of methylation were more prominent compared to gains in DNA methylation patterns, especially in iDMPs exhibiting a constant response.

To explore the biological significance of the genes mapping to the set of iDMPs, we annotated their corresponding gene ontology (GO) term and searched for over-representation of categories using the package GOseq, which weights genes based on number of probes per gene. Enriched biological functions were summarised and visualised in Fig. 4 using semantic similarity, which provides a measure of functional similarity. Functionally similar GO terms appear closer in the plot.

Comparison with publically available transcription data. We compared the iDMPs with transcriptomic data from a previous study of *B. pseudomallei* infection in humans which identified 2604 human genes, that were differentially expressed in patients with septicemic melioidosis[6]. Of these, 76 genes were annotated to the iDMPs we identified in our study. These included genes involved in immune system process (*BCL11B*, *CDKN1C*, *GLMN*, *HLX*, *IL1R2*, *IRF8*, *MAEA*, *MEF2C*, *MR1*, *NBEAL2*, *PRKCH*, *PTGDR2*, *STK3*, *TNFSF8*, *TRIM27*), response to stress (*ADRB2*, *APBB1IP*, *DTNBP1*, *FBXO31*, *MARCH1*, *MSRA*, *PKD2*, *SCARB1*, *ZMYND11*), and inflammatory response (*CD44*, *HDAC4*, *HIF1A*, *IL18RAP*, *TOLLIP*, *IER3*, *NT5E*).

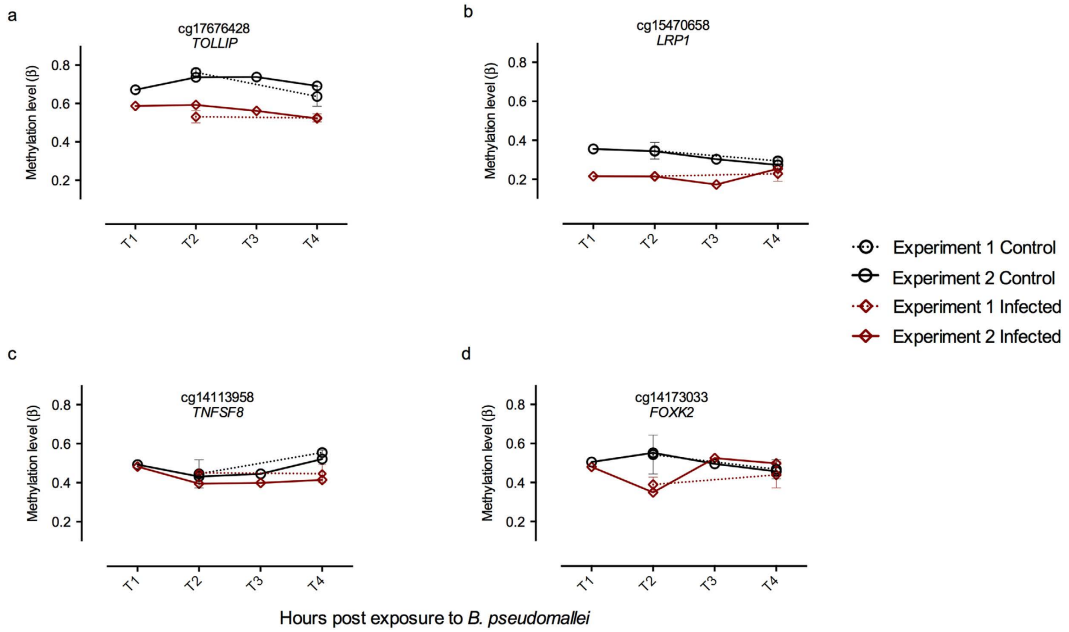


Figure 3. Differential methylation patterns. Each panel shows a representative gene, showing a particular temporal pattern as discussed in the main text and methods. The error bars represent the standard deviation of the methylation levels of the first experiment samples: (**a**) constant hypo-methylation; (**b**) early response hypo-methylation; (**c**) late response hypo-methylation; (**d**) transient response hypo-methylation.

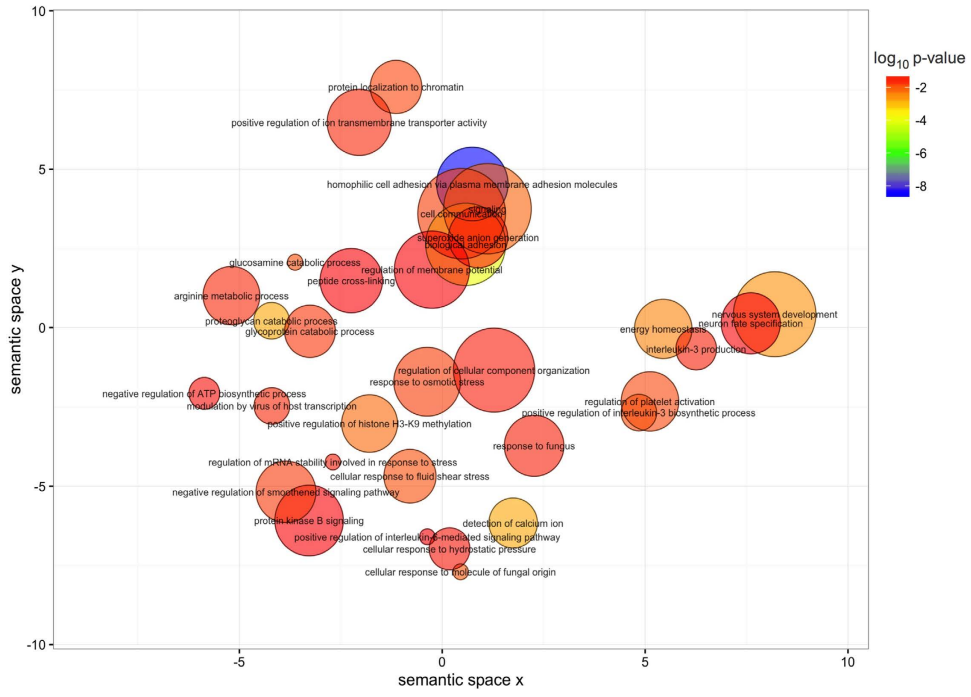


Figure 4. Gene ontology terms enriched ($p < 0.05$) in genes mapping to conserved iDMPs. The colour scale represents the p-values calculated using GOseq. The non-redundant gene ontology terms are clustered using REVIGO.

Comparison with publically available DNA methylation changes during infection with other pathogens. Innate immune cells, such as macrophages or dendritic cells, are recruited in response to pathogens to initiate defense mechanisms. While this study explores DNA methylation changes in macrophages, a recent study identified epigenetic regulation in human dendritic cells before and after *in vitro* infection (at

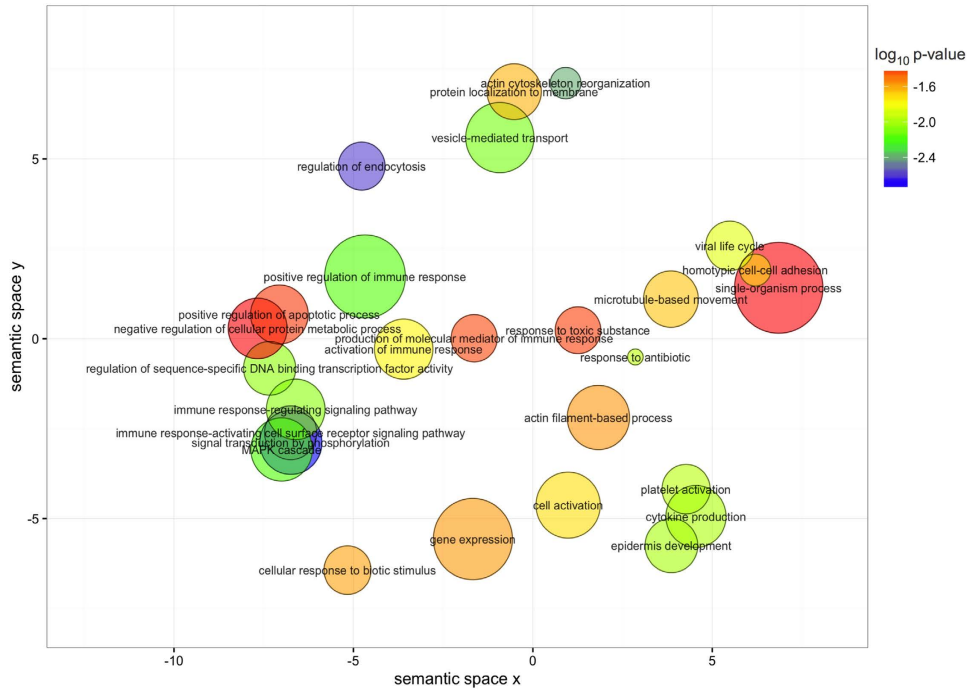


Figure 5. Gene ontology terms enriched ($p < 0.05$) in genes annotated to the iDMPs in *B. pseudomallei* and *M. tuberculosis* infections. The colour scale represents the p-values calculated using GOseq. The non-redundant gene ontology terms are clustered using REVIGO.

18 hours after infection) with *Mycobacterium tuberculosis*[14]. The proportions of iDMPs compared to the total number of sites probed were comparable to our study (0.000912 in our study; 0.000649 in the *M. tuberculosis* study). We did not identify any stringently-filtered iDMPs which were identical in cells infected with *B. pseudomallei* or with *M. tuberculosis* (Supplementary Fig. 4). However, when we considered the genes to which these iDMPs mapped we identified 121 genes (median distance of ~95 kb from the nearest transcription start site), which showed differential methylation in both studies (Supplementary Table 3). These included genes involved in T cell responses (*BCL11B*, *FOXO1*, *KIF13B*, *PAWR*, *SOX4*, *SYK*), actin cytoskeleton organization (*ACTR3*, *CDC42BPA*, *DTNBP1*, *FERMT2*, *PRKCZ*, *RAC1*), and cytokine production (*FOXP1*, *IRF8*, *MR1*) (Fig. 5).

Discussion

This study expands the limited knowledge on the role of epigenetics during infection by reporting temporal changes to the genome-wide methylome of human cells infected with a bacterial pathogen. For our study we have used human macrophages activated with interferon gamma, because we believe that this more accurately reflects the state of macrophages during sepsis caused by *B. pseudomallei*. We found that a range of genes associated with the immune response is differentially methylated in the host after infection with *B. pseudomallei* and these changes were replicated in a second experiment. These methylation events were associated with genes encoding cytokines, chemokines and their receptors (*CCR4*, IL1R2, IL18RAP, *TNFSF15*, *TNFSF8*, *CCL28*) and signalling pathways (*CASP8AP2*, *TOLLIP*, *SYK*, *ZBP1*, *MAP4K4*, *MBIP*). Immune system processes were also enriched in genes near iDMPs identified in this study and genes differentially expressed in patients' blood infected with *B. pseudomallei*. Increased expression of the *TOLLIP* gene has previously been reported in humans infected with *B. pseudomallei*[6]. The precise role of *TOLLIP* in protective immunity is still being clarified. One function of the protein is to interact with *IL1RI*, *TLR2* and *TLR4* after lipopolysaccharide (LPS) activation, thereby modulating the NF-κB and JNK signaling pathways and consequently inflammatory responses to infection[20]. The role of *B. pseudomallei* LPS in virulence[21] and the activation of *TLR2* and *TLR4* signalling during *B. pseudomallei* infection is well documented[22–24]. It is possible that the differential methylation of *TOLLIP* we have seen modulates the early immune response to *B. pseudomallei* LPS.

Another important feature of disease caused by *B. pseudomallei* is the ability of the pathogen to establish chronic or persistent infections. *B. pseudomallei* is an intracellular pathogen, and phagocytes are believed to be an important niche for growth and survival in the host[25,26]. The differential methylation of genes involved in memory T-cell responses, such as *CD44*, *FOXO1* and *FOXC1*, might contribute to the inability of the host to mount responses capable of clearing infection. Other host defence systems involve interferon gamma and nitric oxide. Interferon gamma plays a key role in the control of *B. pseudomallei* infection[27,28] and suppression of the host innate response to *B. pseudomallei* infection has previously been attributed to downregulation of the type I interferon gamma signalling pathway by the bacterial effector TssM[29]. Previously $CD44^{high}$ CD8 T-cells have been shown to be an important source of interferon gamma[30] and we found evidence of differential methylation of *CD44* in our study. We also found evidence of the differential methylation of a number of genes (*IRF-8*, *IFNE* and

ZBP1), which are involved in the regulation of interferon gamma expression. Nitric oxide has potent antibacterial activity towards *B. pseudomallei*[31,32]. A DMP located upstream of the dimethylarginine dimethylaminohydrolase 2 (*DDAH2*) gene that encodes an enzyme in the nitric oxide generation pathway is hypo-methylated at all time points. The differential methylation of the gene encoding the *NOX4 NADPH* oxidase that we have seen, might explain the activation of this enzyme in phagocytes infected by the bacterium[33]. Such observations provide new insights in the possible molecular mechanisms, which underpin suppression of the host response.

We also found differential methylation of genes associated with ubiquitination of proteins targeted for degradation, including the *SPSB4* and *WSB1* gene products, which are associated with substrate recognition and the *NEDD4L* and *SIAH1* genes encoding ubiquitin ligases. The deamidation of the *NEDD8L* protein by a *B. pseudomallei* type III effector protein (*CHBP*) has previously been demonstrated, and this triggers apoptosis of host cells[34]. However, the modulation of host cell ubiquitination has also been associated with the suppression of host immunity[35] and the downregulation of the NF-kappaB/type I interferon signalling pathways has been attributed to the ubiquitination of signalling molecules[29]. Our findings provide additional insight into the mechanisms that contribute to modulation of the ubiquitination pathway.

The most significantly enriched GO terms in the set of iDMPs were cell adhesion related (FDR corrected $P < 0.05$). *B. pseudomallei* is known to promote the polymerisation of host actin at one pole of the bacterium in an *ARP2/3*- dependent manner[36]. This enables motility of the bacteria and is essential for the uptake of bacteria into cells and the spread of bacteria within and between cells. Several iDMPs map to genes involved in actin binding or polymerisation. These include *CDC42BPA*, a component of the BLOC-1 complex, which can interact with *WASH* - a regulator of *ARP2/3*[37] and *ACTR3*, a major component of the *ARP2/3* complex[38]. We also found differential methylation of one of the central regulators of actin polymerisation, the *PAK1* p21-activated kinase[39].

Our findings also suggest that there are some common targets for methylation in cells infected with pathogens. Most of the methylation changes in cells infected with *M. tuberculosis* or *B. pseudomallei* are losses rather than gains in methylation. A broadly similar trend has also been reported by Marr *et al.*[15] who reported a larger proportion of hypo-methylated CpG sites in *L. donovani* infected macrophages. Overall it seems that there is a general trend toward demethylation of host cell DNA during infection. Our comparison of *M. tuberculosis* or *B. pseudomallei* induced changes in methylation patterns revealed that although a number of common genes were differentially methylated, there were no conserved iDMPs. However, since these studies used different cell types we cannot discount the possibility that there would be conserved iDMPs if the experimental conditions were identical.

We also found that the conserved iDMPs were significantly enriched in the first exon of genes, DNA methylation in this genomic compartment has been associated with transcriptional suppression[40,41], suggesting that a large proportion of the infection associated DNA methylation changes observed could be altering gene transcription.

In summary, our work provides new insights into the extent to which infection with a bacterial pathogen results in differential methylation of host cell DNA. Our findings indicate that methylation of host DNA occurs on a much greater scale than previously suggested. None of the effector *B. pseudomallei* molecules identified to date have been shown to methylate host DNA. Whether differential methylation is a consequence of the direct interaction of bacterial methyltransferases with host cell DNA, or the consequence of modification of host cell methyltransferase activity awaits investigation.

Methods

Cell line and infection model. The human leukemic monocyte lymphoma cell line (U937, ATCC CRL-1593.2) was maintained in RPMI 1640 supplemented with 10% foetal bovine serum (FBS) at 37 °C. U937 cells were differentiated to macrophage-like cells following exposure to 20 ng/ml (final concentration) of phorbol 12-myristate 13-acetate (PMA) for 48 hours at 37 °C and differentiation evidenced by increased adherence to tissue culture flasks. A total of eight flasks (four for infected cells and four for uninfected controls) were prepared.

Overnight cultures of *B. pseudomallei* K96243 were diluted in L-15 medium and added to differentiated U937 cells at a multiplicity of infection (MOI) of 10. Where indicated *B. pseudomallei* K96243 expressing red fluorescent protein (RFP) was used[42]. Uninfected controls in the remaining four flasks were overlaid with L15 medium only. The cells were then incubated at 37 °C for 2h to allow infection. The cells were washed 3 times with phosphate buffered saline (PBS) and incubated with fresh L15 medium containing 1 mg/ml kanamycin for 2 hr to kill extracellular bacteria. After 2 hrs the macrophage cells were held in fresh media containing 250 μg/ml kanamycin to supress the growth of extracellular bacteria. At appropriate time points the cells were washed 3 times in warm PBS and lysed with 0.1% (vol/vol) Triton X-100. DNA was isolated using an AllPrep kit (Qiagen) and stored at −80 °C until required. DNA yield was measured using a Nanodrop instrument with measurements between 22.8–50.6 ng/ul.

Enumeration of adhesion and uptake of *B. pseudomallei* by U937 cells. At 1 hour (T1), 2 hours (T2), 4 hours (T4) and 24 hours (T24) post infection the cells were washed 3 times in warm PBS and lysed with 0.1% (vol/vol) triton X-100. Serial dilutions of the cell lysate were plated onto LB agar to determine the intracellular bacterial cell counts (Fig. 2).

Additionally at T2, cells were washed 3 times with PBS and overlaid with 200 μl paraformaldehyde 0.4%, ensuring any coverslips were fully immersed. Cells were than incubated at room temperature for 30 minutes. PFA was removed and coverslips were washed twice with PBS for 1 hour for each wash. Coverslips were removed and stained with 4′,6-diamidino-2-phenylindole (DAPI) and cells visualized with epifluorescence. Eight fields of view

were visualised and the number of uninfected or infected U937 cells counted. The mean number of cells associated with bacteria at T2 was calculated (Fig. 1).

A second experiment was designed to enable verification of the results of the first experiment. Similar procedures were followed as described above and DNA collected from uninfected and infected cells at 1 hour (T1), 2 hours (T2), 3 hours (T3) and 4 hours (T4) post infection.

DNA methylation analysis. The DNA methylation profile of the infected macrophage DNA was determined using the Infinium HumanMethylation450 BeadChip (450K) (Illumina Inc.) following the manufacturer's instructions. The 450 k BeadChip interrogates DNA methylation at >480,000CpG sites across the genome. Briefly, 500 ng of genomic DNA was sodium bisulfite converted using the EZ-96 DNA Methylation kit (Zymo research, CA, USA) using the manufacturer's instructions. Post-hybridisation allelic-specific single-base extension of the probes incorporates a fluorescent label enabling detection. Replicates were processed together on the same array to reduce batch effects. The data was extracted and the initial analysis was performed using GenomeStudio (2010.3) methylation module (1.8.5). For further details please see Pidsley *et al.*[43].

Data Analysis. For each CpG site a beta value was generated by the relative intensity of the green fluorescent signal (methylated (M)) to the red and green signal combined (M/M+U+100). Quality control checks and quantile normalisation were implemented using WateRmelon[43]. Samples with more than 1% of sites with a detection p-value greater than 0.05 were removed as were probes with 1% of samples with a detection p-value greater than 0.05. Probes were removed if they had a bead count less than 3 in 1% of samples. Cross-hybridizing probes were removed[44], leaving 425496 probes for analysis. DMPs between infected and control cells were identified using $\Delta\beta$ cut-off of 0.1. iDMPs identified at T2 and T4 in the first experiment, were compared to the corresponding probes at the same time point in the second experiment. Probes exhibiting a $\Delta\beta$ change in the same direction of $=>0.1$ were taken for further analysis. Correlation of $\Delta\beta$ values between the two experiments was measured using Pearson Correlation Coefficient. This set of replicated iDMPs were tested for enrichment of genomic regions using a two-tailed Fisher's exact test, compared to the frequency of all probes on the Illumina 450 K array. iDMPs with consistent up or down-methylation throughout T1, T2, T3 and T4 were determined. The temporal methylation patterns were assigned to categories based on the change in score at successive time points. As many of the differential methylation patterns found in disease or environmental factors are characterized by smaller changes, in the range of 5%[15], this criterion was selected to identify temporal patterns of differential methylation. Signatures with more than 5% differential methylation at all time points were classified as, constant hyper- (+5%) or constant hypo-methylated (−5%) probes. Signatures with more than 5% differential methylation at early time points (T1, T2, T3) and no change at the T4 time point were classified as early response probes. Signatures with no response at the T1 time point and with more than 5% differential methylation at later time points were classified as late response probes. Signatures with no response at T1 and T4 and more than 5% differential methylation at T2 or T3 were classified as transient responses. The remainder of the patterns were classified as oscillatory probes.

Candidate genes were assigned to the probes using the GREAT software[45], genes are allotted to genomic regions taking into account the functional significance of *cis*-regulatory regions. Gene ontology term enrichment analysis was performed using the Bioconductor package GOseq[46]. Enriched gene ontology terms are selected based on the criteria of having p-value < 0.05. Redundant gene ontology terms were removed and the non-redundant gene ontology terms were clustered using semantic similarity measures (the *simRel* score) using REVIGO[47].

References

1. Limmathurotsakul, D. *et al.* Predicted global distribution of *Burkholderia pseudomallei* and burden of melioidosis. *Nat. Microbiol.* **1,** 15008 (2016).
2. Wiersinga, W. J., van der Poll, T., White, N. J., Day, N. P. & Peacock, S. J. Melioidosis: insights into the pathogenicity of *Burkholderia pseudomallei*. *Nat. Rev. Microbiol.* **4,** 272–282 (2006).
3. Galyov, E. E., Brett, P. J. & DeShazer, D. Molecular insights into *Burkholderia pseudomallei* and *Burkholderia mallei* pathogenesis. *Annu. Rev. Microbiol.* **64,** 495–517 (2010).
4. Stevens, M. *et al.* Identification of a bacterial factor required for actin-based motility of *Burkholderia pseudomallei*. *Mol. Microbiol.* **56,** 40–53 (2005).
5. Chin, C.-Y., Monack, D. M. & Nathan, S. Genome wide transcriptome profiling of a murine acute melioidosis model reveals new insights into how *Burkholderia pseudomallei* overcomes host innate immunity. *BMC Genomics* **11,** 672 (2010).
6. Pankla, R. *et al.* Genomic transcriptional profiling identifies a candidate blood biomarker signature for the diagnosis of septicemic melioidosis. *Genome Biol.* **10,** R127 (2009).
7. Joost Wiersinga, W., Dessing, M. C. & van der Poll, T. Gene-expression profiles in murine melioidosis. *Microbes Infect.* **10,** 868–877 (2008).
8. Takahashi, K. Influence of bacteria on epigenetic gene control. *Cellular and Molecular Life Sciences* **71,** 1045–1054 (2014).
9. Bierne, H., Hamon, M. & Cossart, P. Epigenetics and bacterial infections. *Cold Spring Harbor perspectives in medicine* **2,** a010272 (2012).
10. Suzuki, M. M. & Bird, A. DNA methylation landscapes: provocative insights from epigenomics. *Nat. Rev. Genet.* **9,** 465–476 (2008).
11. Ushijima, T. *et al.* Fidelity of the methylation pattern and its variation in the genome. *Genome Res.* **13,** 868–874 (2003).
12. Huang, F. Y. *et al. Helicobacter pylori* induces promoter methylation of E-cadherin via interleukin-1beta activation of nitric oxide production in gastric cancer cells. *Cancer* **118,** 4969–4980 (2012).
13. Niwa, T. *et al.* Inflammatory processes triggered by *Helicobacter pylori* infection cause aberrant DNA methylation in gastric epithelial cells. *Cancer Res.* **70,** 1430–1440 (2010).
14. Pacis, A. *et al.* Bacterial infection remodels the DNA methylation landscape of human dendritic cells. *Genome Res.* **25,** 1801–1811 (2015).
15. Marr, A. K. *et al. Leishmania donovani* Infection Causes Distinct Epigenetic DNA Methylation Changes in Host Macrophages. *PLoS Pathog.* **10,** e1004419 (2014).

16. Takahashi, K. *et al.* Epigenetic control of the host gene by commensal bacteria in large intestinal epithelial cells. *J. Biol. Chem.* **286,** 35755–35762 (2011).
17. Lehmann, M. H. Recombinant human granulocyte-macrophage colony-stimulating factor triggers interleukin-10 expression in the monocytic cell line U937. *Mol. Immunol.* **35,** 479–485 (1998).
18. Depke, M. *et al.* Bone marrow-derived macrophages from BALB/c and C57BL/6 mice fundamentally differ in their respiratory chain complex proteins, lysosomal enzymes and components of antioxidant stress systems. *J. Proteomics* **103,** 72–86 (2014).
19. Mill, J. & Heijmans, B. T. From promises to practical strategies in epigenetic epidemiology. *Nature reviews. Genetics* **14,** 585–594 (2013).
20. Capelluto, D. G. S. Tollip: A multitasking protein in innate immunity and protein trafficking. *Microbes Infect.* **14,** 140–147 (2012).
21. DeShazer, D., Brett, P. J. & Woods, D. E. The type II O-antigenic polysaccharide moiety of *Burkholderia pseudomallei* lipopolysaccharide is required for serum resistance and virulence. *Mol. Microbiol.* **30,** 1081–1100 (1998).
22. West, T. E. *et al.* Toll-like receptor 4 region genetic variants are associated with susceptibility to melioidosis. *Genes Immun.* **13,** 38–46 (2012).
23. West, T. E., Ernst, R. K., Jansson-Hutson, M. J. & Skerrett, S. J. Activation of Toll-like receptors by *Burkholderia pseudomallei*. *BMC Immunol.* **9,** 46 (2008).
24. Wiersinga, W. J. *et al.* Toll-like receptor 2 impairs host defense in gram-negative sepsis caused by *Burkholderia pseudomallei* (Melioidosis). *PLoS Med.* **4,** 1268–1280 (2007).
25. Limmathurotsakul, D. & Peacock, S. J. Melioidosis: A clinical overview. *British Medical Bulletin* **99,** 125–139 (2011).
26. Wiersinga, W. J., Currie, B. J. & Peacock, S. J. Melioidosis. *N. Engl. J. Med.* **367,** 1035–1044 (2012).
27. Santanirand, P., Harley, V. S., Dance, D. A. B., Drasar, B. S. & Bancroft, G. J. Obligatory role of gamma interferon for host survival in a murine model of infection with *Burkholderia pseudomallei*. *Infect. Immun.* **67,** 3593–3600 (1999).
28. Lauw, F. N. *et al.* Soluble granzymes are released during human endotoxemia and in patients with severe infection due to Gram-negative bacteria. *J. Infect. Dis.* **182,** 206–213 (2000).
29. Tan, K. S. *et al.* Suppression of host innate immune response by *Burkholderia pseudomallei* through the virulence factor TssM. *J. Immunol.* **184,** 5160–5171 (2010).
30. Lertmemongkolchai, G., Cai, G., Christopher, A., Bancroft, G. J. & Hunter, C. A. Bystander Activation of CD8 + T Cells Contributes to the Rapid Production of IFN- γ in Response to Bacterial Pathogens. *J. Immunol.* **166,** 1097–1105 (2013).
31. Jones-Carson, J., Laughlin, J. R., Stewart, A. L., Voskuil, M. I. & Vázquez-Torres, A. Nitric oxide-dependent killing of aerobic, anaerobic and persistent *Burkholderia pseudomallei*. *Nitric Oxide - Biol. Chem.* **27,** 25–31 (2012).
32. Breitbach, K., Wongprompitak, P. & Steinmetz, I. Distinct roles for nitric oxide in resistant C57BL/6 and susceptible BALB/c mice to control *Burkholderia pseudomallei* infection. *BMC Immunol.* **12,** 20 (2011).
33. Wikraiphat, C., Pudla, M., Baral, P., Kitthawee, S. & Utaisincharoen, P. Activation of NADPH oxidase is essential, but not sufficient, in controlling intracellular multiplication of *Burkholderia pseudomallei* in primary human monocytes. *Pathog. Dis.* **71,** 69–72 (2014).
34. Yao, Q. *et al.* Structural mechanism of ubiquitin and NEDD8 deamidation catalyzed by bacterial effectors that induce macrophage-specific apoptosis. *Proc. Natl. Acad. Sci. USA* **109,** 20395–20400 (2012).
35. Lee, S. H. *et al. Burkholderia pseudomallei* suppresses *Caenorhabditis elegans* immunity by specific degradation of a GATA transcription factor. *Proc Natl Acad Sci USA* **110,** 15067–15072 (2013).
36. Stevens, J. M. *et al.* Actin-binding proteins from *Burkholderia mallei* and *Burkholderia thailandensis* can functionally compensate for the actin-based motility defect of a *Burkholderia pseudomallei* bimA mutant. *J. Bacteriol.* **187,** 7857–7862 (2005).
37. Ryder, P. V. *et al.* The WASH complex, an endosomal Arp2/3 activator, interacts with the Hermansky-Pudlak syndrome complex BLOC-1 and its cargo phosphatidylinositol-4-kinase type IIα. *Mol. Biol. Cell* **24,** 2269–2284 (2013).
38. May, R. C. The Arp2/3 complex: a central regulator of the actin cytoskeleton. *Cell Mol Life Sci* **58,** 1607–1626 (2001).
39. Sells, M. a *et al.* Human p21-activated kinase (Pak1) regulates actin organization in mammalian cells. *Curr. Biol.* **7,** 202–210 (1997).
40. Brenet, F. *et al.* DNA methylation of the first exon is tightly linked to transcriptional silencing. *PLoS One* **6,** e14524 (2011).
41. Chuang, T.-J., Chen, F.-C. & Chen, Y.-Z. Position-dependent correlations between DNA methylation and the evolutionary rates of mammalian coding exons. *Proc. Natl. Acad. Sci.* **109,** 15841–15846 (2012).
42. Wand, M. E., Müller, C. M., Titball, R. W. & Michell, S. L. Macrophage and *Galleria mellonella* infection models reflect the virulence of naturally occurring isolates of *B. pseudomallei*, *B. thailandensis* and *B. oklahomensis*. *BMC Microbiol.* **11,** 11 (2011).
43. Pidsley, R. *et al.* A data-driven approach to preprocessing Illumina 450K methylation array data. *BMC Genomics* **14,** 293 (2013).
44. Price, M. E. *et al.* Additional annotation enhances potential for biologically-relevant analysis of the Illumina Infinium HumanMethylation450 BeadChip array. *Epigenetics Chromatin* **6,** 4 (2013).
45. McLean, C. Y. *et al.* GREAT improves functional interpretation of cis-regulatory regions. *Nat. Biotechnol.* **28,** 495–501 (2010).
46. Young, M. D., Wakefield, M. J., Smyth, G. K. & Oshlack, A. Gene ontology analysis for RNA-seq: accounting for selection bias. *Genome Biol.* **11,** R14 (2010).
47. Supek, F., Bosnjak, M., Skunca, N. & Smuc, T. Revigo summarizes and visualizes long lists of gene ontology terms. *PLoS One* **6,** e21800 (2011).

Acknowledgements

This work was supported by the UK Defence Science and Technology Laboratory, award DSTLX-1000060221 to R.W.T., O.S.S. and O.E.A.

Author Contributions

D.C., O.L.C., O.S.S., J.M. and R.W.T. designed the study. O.L.C. and S.W. established the infection model. E.L.D. conducted the DNA methylation experiments. D.C. and E.L.D. analysed the DNA methylation data. D.C., E.L.D., O.L.C., O.E.A., J.L.P., O.S.S., J.M. and R.W.T. analysed the results. D.C. and R.W.T. wrote the manuscript. All the authors read and approved the manuscript.

Additional Information

Accession codes: All of the DNA methylation data are available at the GEO database (accession number GSE83379).

Supplementary information accompanies this paper at http://www.nature.com/srep

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Cizmeci, D. *et al.* Mapping epigenetic changes to the host cell genome induced by *Burkholderia pseudomallei* reveals pathogen-specific and pathogen-generic signatures of infection. *Sci. Rep.* **6**, 30861; doi: 10.1038/srep30861 (2016).